M100 21

THE OPEN UNIVERSITY

Mathematics Foundation Course Unit 21

Probability and Statistics III

The Open University

Mathematics Foundation Course Unit 21

PROBABILITY AND STATISTICS III

Prepared by the Mathematics Foundation Course Team

Correspondence Text 21

The Open University Press

Professor Frank Downton acted as consultant for this unit.

The Open University Press
Walton Hall, Bletchley, Buckinghamshire

First published 1971. Reprinted 1972

Designed by the Media Development Group of the Open University

Printed in Great Britain by
EYRE AND SPOTTISWOODE LIMITED
AT THE GROSVENOR PRESS PORTSMOUTH

SBN 335 01020 2

Open University courses provide a method of study for independent learners through an integrated teaching system, including textual material, radio and television programmes and short residential courses. This text is one of a series that makes up the correspondence element of the Foundation Course.

The Open University's courses represent a new system of university-level education. Much of the teaching material is still in a developmental stage. Courses and course materials are, therefore, kept continually under revision. It is intended to issue regular up-dating notes as and when the need arises, and new editions will be brought out when necessary.

For general availability of supporting material referred to in this book, please write to the Director of Marketing, The Open University, Walton Hall, Bletchley, Buckinghamshire.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Bletchley, Buckinghamshire.

1.2

Contents

Objectives

The general aim of this unit is to introduce random variables, probability distributions, and sampling, all of which are fundamental to statistical work.

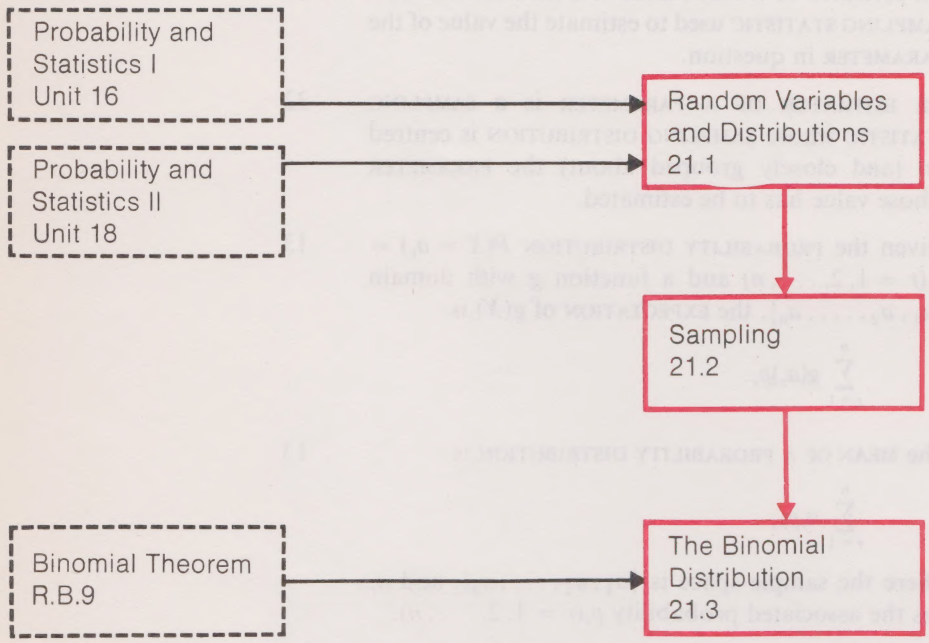
After working through this unit you should be able to:

(i) select the most suitable sample space in any given (simple) situation;
(ii) explain what is meant by a random variable and its probability distribution;
(iii) explain what is meant by the expectation of a random variable, and calculate the expectation in simple cases;
(iv) derive the variance of a simple discrete distribution;
(v) compare two distributions in terms of their means and variances;
(vi) explain what is meant by a sampling distribution, and distinguish it from the parent probability distribution;
(vii) use sampling statistics to estimate unknown parameters, and compare the suitability of different statistics for this purpose in simple cases;
(viii) derive the theoretical form of the binomial distribution as a particular case of a sampling distribution;
(ix) find the mean and variance of the binomial distribution.

Note

Before working through this correspondence text, make sure you have read the general introduction to the mathematics course in the Study Guide, as this explains the philosophy underlying the whole course. You should also be fámiliar with the section which explains how a text is constructed and the meanings attached to the stars and other symbols in the margin, as this will help you to find your way through the text.

Structural Diagram



Glossary

Page

Terms which are defined in this glossary are printed in CAPITALS.

BINOMIAL DISTRIBUTION	The BINOMIAL DISTRIBUTION is the PROBABILITY DISTRIBUTION in which $$P(X = r) = \binom{n}{r} p^r(1-p)^{n-r}$$ $(r = 0, 1, \ldots, n)$, where n and p are PARAMETERS.	36
ESTIMATE OF A PARAMETER	An ESTIMATE OF A PARAMETER is a REALIZATION of a SAMPLING STATISTIC used to estimate the value of the PARAMETER in question.	32
ESTIMATOR OF A PARAMETER	An ESTIMATOR OF A PARAMETER is a SAMPLING STATISTIC whose SAMPLING DISTRIBUTION is centred on (and closely grouped about) the PARAMETER whose value has to be estimated.	32
EXPECTATION	Given the PROBABILITY DISTRIBUTION $P(X = a_r) = p_r (r = 1, 2, \ldots, n)$ and a function g with domain $\{a_1, a_2, \ldots, a_n\}$, the EXPECTATION of $g(X)$ is $$\sum_{r=1}^{n} g(a_r)p_r.$$	13
MEAN OF A PROBABILITY DISTRIBUTION (OR OF A RANDOM VARIABLE)	The MEAN OF A PROBABILITY DISTRIBUTION is $$\sum_{r=1}^{n} a_r p_r,$$ where the sample space is $\{a_1, a_2, \ldots, a_n\}$, and a_r has the associated probability $p_r (r = 1, 2, \ldots, n)$.	13
PARAMETER OF A PROBABILITY DISTRIBUTION	A PARAMETER OF A PROBABILITY DISTRIBUTION is a mathematical parameter (i.e. an arbitrary constant which distinguishes various specific cases) occurring in the expression of the probability distribution.	30
PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE X	A PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE, X, is a function $$a_r \longmapsto p_r (r = 1, 2, \ldots, n).$$ The domain of the function is a numerical sample space $\{a_1, a_2, \ldots, a_n\}$, where the probability associated with a_r is $p_r (r = 1, 2, \ldots, n)$. (By the laws of probability, we have $0 \leqslant p_r \leqslant 1$ $(r = 1,2, \ldots, n)$, and $$\sum_{r=1}^{n} p_r = 1.)$$ We write $$P(X = a_r) = p_r (r = 1, 2, \ldots, n).$$	11
RANDOM SAMPLE FROM A DISTRIBUTION	A RANDOM SAMPLE FROM A DISTRIBUTION is a sequence of REALIZATIONS of a RANDOM VARIABLE, obtained from a sequence of independent trials.	21
RANDOM VARIABLE	A RANDOM VARIABLE is any numerical quantity whose value is determined by the outcome of a trial where the outcome is not predictable.	8
REALIZATION OF A RANDOM VARIABLE, X	A REALIZATION OF A RANDOM VARIABLE, X, is a numerical value taken by X in a particular trial.	8

		Page
SAMPLING DISTRIBUTION	A SAMPLING DISTRIBUTION is the PROBABILITY DISTRIBUTION of a SAMPLING STATISTIC.	27
SAMPLING STATISTIC	A SAMPLING STATISTIC is a RANDOM VARIABLE based on a RANDOM SAMPLE rather than on an outcome of an individual trial.	27
STANDARD DEVIATION OF A DISTRIBUTION (OR OF A RANDOM VARIABLE)	The STANDARD DEVIATION OF A DISTRIBUTION is the positive square root of the VARIANCE OF THE DISTRIBUTION (OR OF THE RANDOM VARIABLE).	17
VARIANCE OF A DISTRIBUTION (OR OF A RANDOM VARIABLE)	The VARIANCE OF A DISTRIBUTION (OR OF A RANDOM VARIABLE) is the EXPECTATION of $g(X)$ where $g : X \longmapsto (X - \mu)^2$ and μ is the MEAN OF THE DISTRIBUTION (OR OF THE RANDOM VARIABLE).	15

Notation

Page

The symbols are presented in the order in which they appear in the text.

X	The usual notation for a random variable.	8
$P(X = a_r)$	The probability that the random variable X takes the value a_r.	8
$E(X)$	The expectation of X.	13
μ	The mean of a distribution (or of a random variable).	13
$E(g(X))$	The expectation of $g(X)$.	14
$\text{var}(X)$	The variance of a distribution (or of a random variable).	15
σ	The standard deviation of a distribution (or of a random variable).	17
$P(X > Y)$	The probability that the random variable X takes a value greater than that of the random variable Y.	19

Bibliography

This correspondence text tries to tell you as much as possible about statistics while covering a minimum of the technical detail. No published book is written in quite this way, so it is difficult to make suggestions.

However,

F. Mosteller, R. E. K. Rourke and G. B. Thomas, *Probability with Statistical Applications* 2nd edition (Addison-Wesley 1961)

and

S. N. Collings, *Theoretical Statistics: Basic Ideas* (Macdonald, to be published in late 1971)

will both help on particular topics such as random variables, discrete distributions, expectations, means and variances, and the binomial distribution. Both books carry the subject beyond the scope of *Unit 21*. Mosteller goes a long way beyond, introducing continuous probability distributions in the process; Collings gives more attention to the sections which are relevant to the unit.

21.0 INTRODUCTION

21.0

Introduction
* *

In *Unit 16, Probability and Statistics I* we considered the nature and manipulation of simple data, and in *Unit 18, Probability and Statistics II* we considered probability and its rules of operation. This unit is concerned with *statistics*: the application of probability theory to physical situations with a view to the interpretation of the data and the drawing of inferences. Such inferences are drawn by means of standard procedures on which most, although not all, statisticians are agreed. Indeed, one way of thinking about statistical theory is that it is the body of such standard procedures. By analogy with this, integration theory would be simply the body of standard integrals. But just as there is more to integration theory than the mechanical evaluation of integrals, so there is more to statistical theory than the rote churning out of inferences. Without some understanding of the basic difficulties, the advantages and disadvantages of the various procedures and the inherent assumptions, it is impossible to exercise any discretionary choice between possible procedures or to know to what extent apparent conclusions need to be qualified. Such understandings form an important part of statistical theory — a part which is seldom the subject matter of a textbook. Most statistical books concentrate on the standard procedures although, admittedly, in the end they convey something of the background problems. We shall not have time to give an adequate coverage of standard procedures in this unit, so we shall not attempt to do so; in any case, for positive reasons, we prefer to go into the background problems. In the course of this we shall mention some standard procedures, but only to illustrate points in the argument.

To reinforce the distinction between standard procedures and background understanding, let us take an analogy from the clothing industry. A salesman who hands down a ready-made suit to his client must know his job first; but he clearly knows far less about clothes than a tailor who is capable of making a suit to measure. A tailor will know at least as much as a salesman about what is the most suitable cloth for the purpose; in addition, he will know how the seams hold together, and he will be able to carry out those adjustments which make all the difference between a good fit and a poor fit.

Our objective is not to present an exhaustive thesis on background problems, but to show that there *are* background problems. In a sense, therefore, we are raising more problems than we are solving. In this, what we are doing is akin to philosophy: we are identifying the logical and mathematical problems raised in arguing from particular data to general properties. As in many areas of philosophy, the solutions to these problems are often controversial. Because we are discussing general principles, we shall not make a point of varying the physical settings; on the contrary, we shall stay on familiar ground. We shall mention dice, and dice again; but each time there will be something different, some new mental attitude to assimilate.

We begin by looking again at mathematical models, or rather, in this case, *probability* models. We have already met the urn model (*Unit 18*), in which drawing balls randomly from an urn is equivalent to selecting people randomly, or throwing a fair die. Now, given balls in an urn, and given a random selection procedure as discussed in *Unit 18*, we can work out the probability of any outcome or combination of outcomes; that is, we can apply probability theory. In practice, however, we have people; we have dice; we have processes for selecting the people and for throwing the dice. But we cannot know with certainty that the physical operations we carry out on physical objects are accurately represented by the urn model. To pursue this further, let us restrict ourselves to the case of a single die.

A die is said to be *fair* if and only if: for a single throw, the sample space is $\{1, 2, 3, 4, 5, 6\}$, where

(i) the probability of each elementary event is $\frac{1}{6}$;
(ii) the elementary events are statistically independent.

In the case of a fair die, the urn model with six numbered balls, which are replaced after selection, faithfully represents the situation, and there is no problem. But *is* the particular die fair? In other words, *is* the model accurate for this die? The only way to find out is to experiment and see, and the first thing we realize is that it is in fact quite difficult to detect bias. Suppose that instead of being fair the die were biassed towards the 6 in such a way that the probability of obtaining a 6 was $\frac{1}{4}$ and of obtaining each of the other numbers was $\frac{3}{20}$. This is a fairly substantial bias towards the 6, representing an increase of 50 per cent in the probability of a 6 occurring. Yet to obtain sufficient data to discriminate between a fair die and one with this bias, with only a small risk of making the wrong decision, would require the die to be thrown several hundred times.

In most situations where the statistician is attempting to interpret data, a validation of the model is not possible. This is why in most situations a statistician has no alternative but to use his judgment and experience to construct a model, and his interpretation of data will always contain the explicit or implicit statement "if the model is correct". Arguments about the interpretation of statistics are thus almost invariably not about the computations (which may easily be seen to be right or wrong) but about whether the model on which the interpretation is based is valid or not. Because statistical investigations often deal with questions which give rise to strong emotions (for example, the long drawn out controversy about whether smoking causes lung cancer), the real nature of the argument is commonly lost. Also, because the validation of a model, even when sufficient information is available, is usually a sophisticated statistical exercise, the statistician looks at the implications which simple models have for the interpretation of data; this is the attitude adopted in this unit. In practical situations the experienced statistician may well have reservations about his interpretation because of the limitations of his model. This does not necessarily prevent him from using that model, which may be the only way of making *any* interpretation.

To see what is meant by this last remark, suppose we doubt the accuracy of the urn model for representing a single die because the event $\{6\}$ appears to have probability greater than $\frac{1}{6}$. If we do not know the true probability, we have to estimate it. We might throw the die 1 000 times and get a 6 on 254 occasions. In this case we would accept $\frac{1}{4}$ as the probability of getting a 6, and you might think we had made no questionable assumptions. We have not taken any model of the situation; we have simply carried out an experiment with a completely open mind, and drawn our conclusions from the data. But despite first appearances, we have made assumptions! We have assumed (for instance) that

(i) the probability of a 6 is constant; that is, independent of the *number* of earlier throws of the die;

(ii) the event $\{6\}$ is statistically independent of the other elementary events; that is, the probability of a 6 is independent of the *results* of earlier throws.

The point is that these are assumptions, and they could be invalid: (i) could be false if the die suffered appreciably from wear, and (ii) could be false if the table were slightly damp so that the face last in contact with the surface was rather stickier than the other faces. Yet unless we make these kinds of assumption, we can scarcely begin to analyse the data and draw any conclusions.

Thus we cannot avoid making assumptions and constructing models. That is why one has to acquire a reasonable "feel" for the subject, to learn by experience which models are acceptable, and which are not, in any given situation.

Let us, however, turn from these general considerations to particular situations and ideas.

21.1 RANDOM VARIABLES AND DISTRIBUTIONS

21.1

21.1.0 Introduction

21.1.0
Introduction

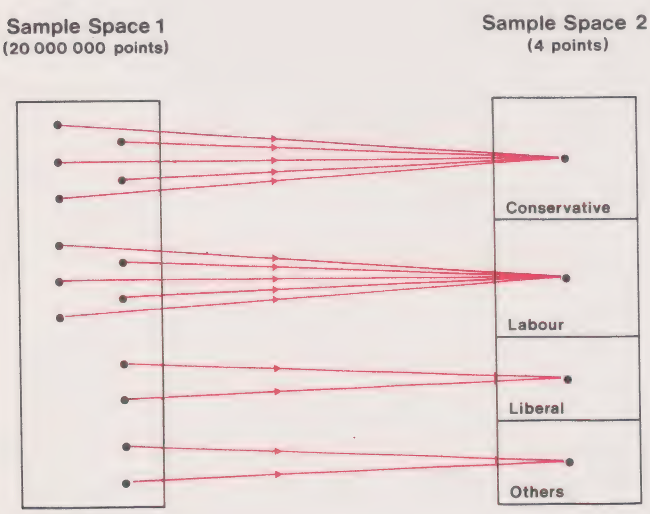
In *Unit 18, Probability and Statistics II*, we saw that the idea of a sample space whose elements (points) represent all the possible outcomes of a trial is fundamental to a study of probability and statistics. However, in probability and statistics, as in other branches of mathematics, the simple ideas often only enable us to deal with simple problems. To do more than this requires more powerful mathematical tools, which may seem at first sight to take us further away from the physical problems with which they are concerned. Before beginning to develop these more powerful methods, it is worth while looking at a particular example to see why the simple probability concepts are unlikely to be adequate to cope with real problems.

21.1.1 The Choice of a Sample Space

21.1.1

Discussion

Suppose a population of 20 000 000 voters contains 9 000 000 (45 per cent) who, if they vote, support the Conservative Party, 9 000 000 (also 45 per cent) who support the Labour Party, 1 000 000 (5 per cent) who support the Liberal Party and 1 000 000 (5 per cent) who support other parties. Even without any more than an intuitive idea of the nature of probability, most people would be happy with the statement that the probability of a "man in the street" being a supporter of the Conservative Party is 0.45. As outlined in *Unit 18*, we could imagine a sample space of 20 000 000 points all with equal probability, each point representing a particular voter. The event "a man in the street supports the Conservative Party" then consists of all the 9 000 000 points representing Conservative supporters, and the probability of this event is given by $9 \times 10^6/20 \times 10^6 = 0.45$. The advantage of making this explicit is that it lays bare the assumption implied by the phrase "a man in the street". We do not mean by this phrase a voter approached, say, by a television interviewer at a street corner, but a voter selected in some scientific way, *so that all voters are equally likely to be chosen*. Assuming that the selection of the "man in the street" has been made in this scientific manner, the way in which the sample space is defined is not unique. If our only purpose is to record the probabilities of selecting the four different kinds of voter, it is sufficient to consider a sample space containing only four points labelled *Conservative, Labour, Liberal* and *Others*, with associated probabilities 0.45, 0.45, 0.05 and 0.05 respectively.



Thus, points in the first space are mapped to points in the second, and any alternative descriptions of the results of a particular experiment will be connected with the basic sample space in a similar way. For example, if we were interested only in Conservatives and non-Conservatives, a sample space of two points would be sufficient. The sample space appropriate to a particular trial depends not only on the trial but also on its purpose. These ideas of possible alternative sample spaces, and of the sample space which is appropriate to a particular situation, are developments of the subject beyond the elementary definition given in *Unit 18*. There, a sample space was simply the set of all possible outcomes.

The position becomes more complicated if we want to predict the result of an election in which only three-quarters of the eligible voters participate. Intuitive ideas of probability are insufficient to give any meaningful prediction, even if it is assumed by generalizing from "the man in the street" that all possible groups of 15 000 000 voters are equally likely to exercise their right to vote. If all voters were treated as individuals and each sample point from this experiment represented a different selection of 15 000 000 voters, the number of points in the resulting sample space would be of the order of $10^{5\,000\,000}$. Clearly such a space needs to be contracted, or at least simplified in some way, for it to become intelligible. The sample space which is likely to lead to a manageable description of the probability situation is not immediately obvious, and in any case it is not unique. The point to be clear about is that it is usually not possible to separate the construction of a sample space from the purpose to which it is to be put.

As another example, a large aircraft might have two tyres on each leg of its undercarriage. Suppose that on landing one (or both) of these tyres might burst. Failure might be caused either by a manufacturing fault or by particularly severe wear due to the condition of the runway. Denoting the two tyres by L (Left) and R (Right), the state of a tyre after landing by F (Failed) and N (Not failed), and the cause of failure by M (Manufacturing fault) and W (Wear), the possible 9 outcomes of an experiment (the aircraft landing) are

(LFM ; RFM)	(LFM ; RFW)	(LFM ; RN)
(LFW ; RFM)	(LFW ; RFW)	(LFW ; RN)
(LN ; RFM)	(LN ; RFW)	(LN ; RN)

However, for the pilot of the aircraft, a sufficient division of the possible results might be a sample space containing the three points:

(i) neither tyre bursts;
(ii) exactly one tyre bursts;
(iii) both tyres burst.

On the other hand, to the tyre manufacturer, a two point sample space might be sufficient, consisting of

(i) no manufacturing faults;
(ii) one or more manufacturing faults.

(In the second case he might be liable for damages.)

To see the sort of thinking needed, try the following exercise.

Exercise 1

Exercise 1
(4 minutes)

When a lift in a block of flats with nine floors is stationary, its doors may be open or closed.

(i) If, at a particular time, the lift is known to be stationary at one of the floors, how many points are there in the sample space which defines the exact state of the lift? (The "state" of the lift means the floor which it is at, and whether the doors are open or closed.)

(ii) If the lift is stationary at some floor and not working, the maintenance man has to carry his tools upstairs from the ground floor to the floor on which the lift is stuck. What sample space is sufficient for his purpose?

(iii) If it takes one second for the lift doors to open or close and one second for the lift to travel between any two adjacent floors, what is the sample space which would suitably describe the state of the lift for a man who presses the CALL button on the middle floor (in the sense of how many seconds it will be before he can step into it)? (Assume that no one else calls the lift.)

(HINT: Find the time corresponding to each of the possible states.) ■

Discussion
* *

These considerations become more important when probability is not used simply as a descriptive model for a physical situation, but is the basis for statistical methods. To return to the 20 000 000 potential voters; for the statistician, the picture of voting behaviour described so far is merely a preliminary discussion of the real problem.

Although it is not the purpose of this unit to go deeply into the solution of problems of voting behaviour, it is worth looking at the real nature of the problem, to see why the simple concept of the sample space with its associated probabilities is insufficient to deal with what the statistician has to do.

The statistician will not, as we previously assumed, know the true proportions of different types of voters in the population. If he is both lucky and industrious, the sort of information available will be:

(i) the total size of the voting population by constituencies (from electoral rolls), together with voting patterns at previous elections;

(ii) the way in which a few thousand people *say* they will vote (from public opinion polls);

(iii) the number of people who are likely to vote and how these people are distributed between the parties (from past experience and opinion polls).

The information in each of the three categories will contain uncertainties which may or may not be expressible in terms of probabilities. To try to describe the overall situation in terms of a single sample space would, even if it were possible, result in a description so complicated as to be useless for predictive purposes. The need to contract this kind of sample space into a manageable pattern is behind the idea of a random variable, which we shall now discuss.

21.1.2 Random Variables

21.1.2

Discussion

Before we give a definition of a random variable, it is worth looking at another example which illustrates in a simple way one of the main purposes of our definition.

Example 1

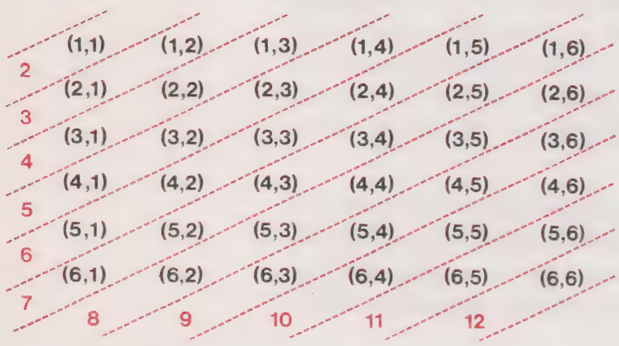
Example 1

Suppose we have an ordinary die whose faces are numbered from 1 to 6. If this die is thrown twice and the number on the uppermost face is recorded for the two throws (as an ordered pair), a sample space containing 36 points will represent the set of possible results. Thus the sample space may be represented as follows:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

If it is assumed that the die is fair, then each of these 36 points has probability $\frac{1}{36}$.

Suppose that we are interested not in the individual results, but in the possible *total* scores obtained from the two throws of this die.



The total scores run from 2 to 12, and the 36 point sample space may (as indicated by the diagonal lines in the array above) be mapped to a new space whose points are the eleven numbers:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Even if we use the model of the fair die, the points in this new space will no longer have equal probability. A total of 2 can be obtained in only one way, as (1, 1), and hence has probability $\frac{1}{36}$. On the other hand, a total of 7 may be obtained in six ways, as (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) and (6, 1), and so the probability of obtaining a total of 7 is the sum of the probabilities of these separate results, which is $\frac{6}{36} = \frac{1}{6}$. ■

The ideas illustrated in this example lead to the definition of a *random variable*.

(*continued on page 8*)

Solution 21.1.1.1

Solution 21.1.1.1

(i) Denoting the floor by a number from 1 to 9, door open by o, and door closed by c, there are 18 points: (1, o), (1, c), (2, o), (2, c), . . . , (9, o), (9, c).
(ii) Since only the floor is relevant, this sample space contains 9 points: 1, 2, . . . , 9.
(iii) The 18 points lead to the corresponding times:

(1, o) 6	(2, o) 5	(3, o) 4	(4, o) 3	(5, o) 0	(6, o) 3	(7, o) 4	(8, o) 5	(9, o) 6
(1, c) 5	(2, c) 4	(3, c) 3	(4, c) 2	(5, c) 1	(6, c) 2	(7, c) 3	(8, c) 4	(9, c) 5

The state of the lift may be described by 7 points representing times of 0, 1, 2, . . . , 6 seconds. ■

(*continued from page 7*)

Consider a trial whose possible outcomes are numbers $a_1, a_2, \ldots, a_n$. We denote by X a general element from this set of numbers, and X is called a random variable. Note that we define a random variable as a general element from a sample space of *numbers*. That is not to say that all other cases are "beyond the pale". There may be some reason for mapping elements of a non-numerical sample space to numerical values. For example, in a particular card game we may not be interested in the suit of a card, and so we can map the sample space of 52 cards to the set $\{1, 2, 3, \ldots, 13\}$; we could then think of a random variable taking values from this new sample space.

Definition 1
* * *

Suppose we have thrown a die three times, and are about to throw it a fourth time. The values on the first three occasions are 2, 5, 4, and on the fourth occasion the value is as yet unknown. We can represent this unknown value by the random variable X which may take any value in the set $\{1, 2, 3, 4, 5, 6\}$.

If the fourth throw gives the value 3, then we say that 3 is the realization of the random variable on this occasion (or that X takes the value 3). In the same way we would say that X had realizations (or took the values) 2, 5, 4 on the first three throws. All we have to keep clear is the distinction between the random variable X when X stands for an unknown value,
and
the realization of X, which is a specific number.

Definition 2
* * *

Before the fourth throw, we are interested in the values X may take on this occasion; but we are also interested in the probabilities with which X may take these values. Thus, for a fair die, X may take the value 1, and the probability that it does so is $\frac{1}{6}$. In a more general situation, the set of possible outcomes of a trial may be the set of numbers $\{a_1, a_2, \ldots, a_n\}$. If the outcome a_r has probability p_r, and X stands for the outcome of a trial, then we denote the probability that X has the realization a_r by $P(X = a_r)$; that is, we write

Notation 1
* * *

$$P(X = a_r) = p_r$$

As we have mentioned, sometimes the outcomes of trials which are not themselves numbers have numbers associated with them; in other words, there is a mapping (a function, in fact) from the sample space to the real numbers R. In this case it is as if the associated real numbers were the outcomes of the trials, so that these numbers can be represented by a random variable. One has to take a little care when this mapping is not one-one, for, in general, if $E_1, E_2, \ldots, E_r$ are all the outcomes mapping to $a (a \in R)$, and X is the random variable standing for the number corresponding to any trial, then

$$P(X = a) = P(E_1) + P(E_2) + \cdots + P(E_r)$$

For example, if the trial consists of drawing a card from a pack, and the aces and court cards are assigned the value 10 whilst every other card is assigned its face value, then

$$P(X = 10) = P(\{\text{Ace}\}) + P(\{\text{King}\}) + P(\{\text{Queen}\}) + P(\{\text{Knave}\}) + P(\{\text{Ten}\})$$
$$= \tfrac{20}{52}$$

Another type of situation in which the outcomes are not single numbers arises when the outcomes are pairs of numbers. For example, if a trial consists of throwing a die twice, the outcomes are ordered pairs of integers (u, v), where $1 \leqslant u \leqslant 6$, and $1 \leqslant v \leqslant 6$. If we are interested simply in the total score, the obvious* thing to do is to map (u, v) to $a = u + v$, and then to take the random variable X having as its realizations all the possible values of a.

Let us see how this works out for a fair die. Only (1, 1) maps to 2. Only (1, 2), (2, 1) map to 3. Only (1, 3), (2, 2), (3, 1) map to 4, etc. In our model for fair dice, each outcome has probability $\frac{1}{36}$, so that we have:

a	2	3	4	5	6	7	8	9	10	11	12
$P(X = a)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Exercise 1

Exercise 1
(3 minutes)

For a fair die which has been thrown twice, give the set, S, of possible values for the appropriate random variable, together with the probabilities associated with these values, if we are interested in:

(i) the larger of the two scores (or the common score if the two are equal);
(ii) the difference between the first score and the second score: (first score) − (second score). ■

Exercise 2

Exercise 2
(4 minutes)

In Exercise 21.1.1.1, the time between a man pressing the CALL button on the middle floor and his stepping into the lift is a random variable. If the lift is equally likely to be at any of the nine floors, and the door is always twice as likely to be closed as open, determine the probability of each of the possible values of the random variable. ■

* "Do not imagine that mathematics is hard and crabbed, and repulsive to common sense. It is merely the etherealization of common sense."
W. Thomson (Lord Kelvin)
S. P. Thompson, *Life of Lord Kelvin* (London, 1910)

Solution 1

Solution 1

From the table given on page 7, we have:

(i) If X is the larger of the two scores, then $S = \{1, 2, 3, 4, 5, 6\}$, and

$$P(X = 1) = \tfrac{1}{36}, \quad P(X = 2) = \tfrac{3}{36},$$
$$P(X = 3) = \tfrac{5}{36}, \quad P(X = 4) = \tfrac{7}{36},$$
$$P(X = 5) = \tfrac{9}{36}, \quad P(X = 6) = \tfrac{11}{36}.$$

(ii) If X is the difference between the first and the second scores, then $S = \{-5, 5, -4, 4, -3, 3, -2, 2, -1, 1, 0\}$, and

$$P(X = -5) = \tfrac{1}{36}, \quad P(X = 5) = \tfrac{1}{36},$$
$$P(X = -4) = \tfrac{2}{36}, \quad P(X = 4) = \tfrac{2}{36},$$
$$P(X = -3) = \tfrac{3}{36}, \quad P(X = 3) = \tfrac{3}{36},$$
$$P(X = -2) = \tfrac{4}{36}, \quad P(X = 2) = \tfrac{4}{36},$$
$$P(X = -1) = \tfrac{5}{36}, \quad P(X = 1) = \tfrac{5}{36},$$
$$P(X = 0) = \tfrac{6}{36}.$$

■

Solution 2

Solution 2

The probability that the lift is on any floor is $\frac{1}{9}$, and, since the door is twice as likely to be closed as open, the probability that it is open on any floor is $\frac{1}{27}$, and the probability that it is closed is $\frac{2}{27}$. Hence, using the results tabulated in Solution 21.1.1.1(iii), together with these probabilities, we have:

Time	0	1	2	3	4	5	6
Probability	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{6}{27}$	$\frac{6}{27}$	$\frac{6}{27}$	$\frac{2}{27}$

■

21.1.3 Probability Distributions

21.1.3

Main Text

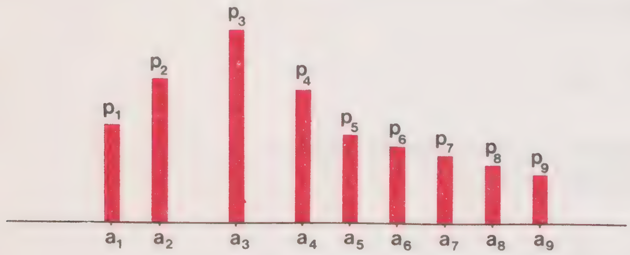
Given a trial with associated numerical outcomes $a_1, a_2, \ldots, a_n$, we have introduced the random variable X and written

$$P(X = a_r) = p_r \qquad (r = 1, 2, \ldots, n),$$

where p_r is the probability of the outcome a_r. Once we know the a's and the p's for any trial, we know how the total probability, namely 1, is *distributed* over the outcomes $a_1, a_2, \ldots, a_n$. In more formal terms, when we have the a's and the p's, we can define a function which maps a_r to its probability p_r; the function has domain the set $\{a_1, a_2, \ldots, a_n\}$ and codomain the set of real numbers R. This function is called the probability distribution of the random variable X. (In fact, it would be better to regard both the set $\{a_1, \ldots, a_n\}$ and the set of images $\{p_1, p_2, \ldots, p_n\}$ as *sequences* rather than *sets*, because we are interested in the *distribution* of the p's, so that each of the p's, whether repeated or not, is of interest.)

Definition 1

The graph of such a probability distribution can be represented by a bar chart; for example:



Starting with a trial having associated numerical outcomes, we are led to consider a random variable X, and then the probability distribution of X, which is specified in terms of numbers a_r and p_r. Conversely, if we take any values a_r $(r = 1, 2, \ldots, n)$ and values p_r satisfying

$$0 \leqslant p_r \leqslant 1$$

and

$$\sum_{r=1}^{n} p_r = 1,$$

there is a conceivable random variable X satisfying

$$P(X = a_r) = p_r.$$

Given any trial with associated numerical outcomes, the first step is usually to work out the probability distribution of the random variable X standing for the outcome; this is either because we simply want the probabilities of the individual outcomes, or because the probability distribution is a half-way stage towards finding something more complicated.

Exercise 1

Exercise 1
(4 minutes)

In a road safety competition the promoters list ten motor accessories (e.g. seat belts, fog lamps) which contribute to road safety. Competitors are asked to select the three which are most effective, and this selection is also made by a panel of experts. A competitor chooses his three items by selecting three balls from an urn containing ten balls labelled with the names of the accessories. If X is the number of accessories which are common to this competitor's choice and that of the experts, find the probability distribution of X.

(HINTS: What is the numerical sample space? How many selections of 3 accessories can be made out of 10? How many of these have 0, 1, 2, 3 accessories from the experts' list?) ■

Solution 1

Solution 1

As suggested in the last question in the hints, the numerical sample space is $\{0, 1, 2, 3\}$. In this exercise we can use as our model an urn containing three white and seven black balls, the three white balls representing the three chosen accessories. Three balls are chosen from this urn in such a way that all possible selections are equally likely. If, for convenience, we regard the balls as distinguishable, we have

$$W_1, W_2, W_3, B_1, B_2, B_3, B_4, B_5, B_6, B_7.$$

The number of selections of 3 balls from these 10 is the number of combinations of 3 objects from 10 objects; that is, $\binom{10}{3} = 120$, all of which are equally likely. The number of such selections containing:

(i) no white balls = number of ways of selecting 3 balls from the 7 black; that is, $\binom{7}{3} = 35$.

(ii) 1 white ball = number of ways of selecting 1 ball from the 3 white and 2 from the 7 black; that is,

$$\binom{3}{1} \times \binom{7}{2} = 3 \times 21 = 63.$$

(iii) 2 white balls = number of ways of selecting 2 balls from the 3 white and 1 ball from the 7 black; that is,

$$\binom{3}{2} \times \binom{7}{1} = 3 \times 7 = 21.$$

(iv) 3 white balls is clearly just 1.

Therefore, we have

$$P(X = 0) = \tfrac{35}{120} = \tfrac{7}{24}$$

$$P(X = 1) = \tfrac{63}{120} = \tfrac{21}{40}$$

$$P(X = 2) = \tfrac{21}{120} = \tfrac{7}{40}$$

$$P(X = 3) = \tfrac{1}{120},$$

and so the probability distribution is the function:

$$0 \longmapsto \tfrac{7}{24}$$

$$1 \longmapsto \tfrac{21}{40}$$

$$2 \longmapsto \tfrac{7}{40}$$

$$3 \longmapsto \tfrac{1}{120}$$

■

21.1.4 Simple Properties of Probability Distributions

21.1.4

Introduction

If a trial is carried out a number of times and the outcome of each trial recorded, then we shall end up with a list of elements chosen from the sample space. If the sample space consists of numbers, then this list will be simply a list of numbers. This list of numbers can be regarded as data to be analysed in much the same way as we analysed data in *Unit 16, Probability and Statistics I*; we can calculate the mean, variance, and so on. By considering these measures when the trial is carried out a large number of times, we are led to define measures which are called the *mean of the distribution*, the *variance of the distribution* and so on. In this section we begin to explain these ideas.

Expectations

Main Text

Suppose X is a random variable with associated probability distribution defined by:

$$P(X = a_r) = p_r \qquad (r = 1, 2, \dots, n).$$

If we were to perform a sequence of N trials and find that X takes the value a_1 on m_1 occasions, a_2 on m_2 occasions, etc., then the average of the values taken by X so far would be

$$\frac{m_1 a_1 + m_2 a_2 + \cdots + m_n a_n}{N}$$

$$= a_1\frac{m_1}{N} + a_2\frac{m_2}{N} + \cdots + a_n\frac{m_n}{N}$$

Now $\frac{m_1}{N}$ is the relative frequency of occurrence of the value a_1, and we have associated the probability of any event with the relative frequency of occurrence of that event in the long run. Hence we would expect $\frac{m_1}{N}$ to be approximately equal to p_1. Similarly we would expect $\frac{m_2}{N}$ to be approximately equal to p_2, and so on. Therefore we would expect the average of the values taken by X to be approximately equal to

$$a_1 p_1 + a_2 p_2 + \cdots + a_n p_n = \sum_{r=1}^{n} a_r p_r$$

Note that N does not appear in this expression. We call this expression the expectation of X, or sometimes the expected value of X; we denote it by $E(X)$ or by the Greek letter μ ("mu"). Thus we have:

Definition 1

Notation 1

$$E(X) = \sum_{r=1}^{n} a_r p_r = \mu$$

μ is often also called the mean of the distribution. Another way of looking at μ is to say that it is the *weighted mean* of the a_r's, the weight of a_r being the probability with which it occurs.

Definition 2

Exercise 1

Exercise 1
(3 minutes)

For the road safety competition of Exercise 21.1.3.1, find the expected number of accessories on which the competitor and the experts agree. ■

Exercise 2

Exercise 2
(2 minutes)

Using the probability distribution for the number of seconds taken by a lift to arrive, already obtained in Exercise 21.1.2.2, determine the expected value of the waiting time. ■

Solution 1

Solution 1

Using the results of Exercise 21.1.3.1,

$$E(X) = \sum_{r=0}^{3} r \times P(X = r)$$

$$= 0 \times \tfrac{7}{24} + 1 \times \tfrac{21}{40} + 2 \times \tfrac{7}{40} + 3 \times \tfrac{1}{120}$$

$$= \frac{21+14+1}{40} = \frac{9}{10}$$ ■

Solution 2

Solution 2

Using the results of Exercise 21.1.2.2,

$$E(X) = \tfrac{2}{27} + \tfrac{8}{27} + \tfrac{18}{27} + \tfrac{24}{27} + \tfrac{30}{27} + \tfrac{12}{27}$$

$$= \tfrac{94}{27}$$

So, on average, one might expect to wait $3\frac{1}{2}$ seconds for the lift. ■

Discussion
* *

Although $E(X)$ is called the expected value of X, we do not *expect* a single random X to take this value, or even be very close to it. It is rare for a random X to equal μ, sometimes impossible, as in the last two exercises. All that we do expect is that with a large number of realizations of X from the same distribution, their average value will be close to μ. For a fair die, the expected value of the random variable is

$$\sum_{r=1}^{6} a_r p_r = \sum_{r=1}^{6} r \times \tfrac{1}{6}$$

$$= \tfrac{21}{6} = 3.5$$

To know the mean of a distribution is to know something, but it is a long way from knowing the whole distribution. A die is not necessarily fair just because the expected value of the score is 3.5. For example, a die whose score is represented by the random variable X, where

$$P(X = 1) = P(X = 6) = \tfrac{1}{4}$$

and

$$P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = \tfrac{1}{8}$$

has expected score 3.5, but it is decidedly "loaded". To achieve a fuller description of the probability distribution we need some other characteristics, and we can get these from a wider application of the idea of an expected value.

Main Text
* *

We introduced expectations by talking about the expected value of X; but there is no reason why we should not consider the expected value of X^2, or indeed of $g(X)$, where g is some general function with domain $\{a_1, a_2, \ldots, a_n\}$ and codomain R. The expectation of X is by definition given by

$$E(X) = \sum_{r=1}^{n} a_r p_r$$

By analogy, the expectation of $g(X)$ is given by

$$E(g(X)) = \sum_{r=1}^{n} g(a_r) p_r,$$

because if $a_r \longmapsto g(a_r)$ and a_r occurs with probability p_r, then $g(a_r)$ occurs with probability p_r. (We need not worry if $g(a_s) = g(a_r)$, as this is looked after by the probability p_s.)

Let us first consider $E(X^2)$. If X is the random variable corresponding to the score obtained with a single throw of a fair die, the expected value of X^2 is

$$E(X^2) = \sum_{r=1}^{6} r^2 p_r = \sum_{r=1}^{6} r^2 \times \tfrac{1}{6} = \tfrac{91}{6}$$

Clearly, for any random variable X there is no end to the possible functions g, but the functions which have either some intuitive physical interpretation or some mathematical or statistical utility are relatively few. One such function enables us to calculate the variance* of a distribution. Consider g, with domain $\{a_1, a_2, \ldots, a_n\}$, where

$$g(X) = (X - \mu)^2, \text{ and } \mu = E(X).$$

The expected value of $(X - \mu)^2$, namely

$$E((X - \mu)^2) = \sum_{r=1}^{n} (a_r - \mu)^2 p_r$$

is called the variance of X, and is denoted by all sorts of things in the literature: notations currently in use are var(X), σ_x^2, σ^2 and μ_2 (σ is the Greek letter "sigma").

Definition 3
* * *

From a computational point of view, it is often easier to make use of the result

$$E((X - \mu)^2) = E(X^2) - \mu^2$$

(see the following exercise).

Exercise 3

Exercise 3
(4 minutes)

(i) Prove that

$$E((X - \mu)^2) = E(X^2) - \mu^2$$

(ii) If X is a random variable standing for the outcome of a fair die, find the variance of X.

(iii) If the die is then loaded so that

$$P(X = 1) = P(X = 6) = \tfrac{1}{4}$$

$$P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = \tfrac{1}{8},$$

find the expected value (μ) of X and the variance of X.

(iv) If $Y = cX + d$ where c, d are constants, find an expression for $E(Y) = \mu_Y$ in terms of $E(X) = \mu_X$.

(v) In (iv), prove that the variance of Y is the product of c^2 and the variance of X. ■

Main Text
* *

We are now going to examine what interpretation (if any) can be put on var (X). Suppose, as we did when discussing expected values, that X is a random variable with probability distribution defined by:

$$P(X = a_r) = p_r \qquad (r = 1, 2, \ldots, n)$$

If we again perform a sequence of N trials and find that, of the outcomes,

m_1 values equal a_1,

m_2 values equal a_2,

$\vdots \quad \vdots \quad \vdots \quad \vdots$

m_n values equal a_n,

then, as before, we would expect $\dfrac{m_r}{N}$ to be approximately equal to p_r.

* See section 16.1.2 of *Unit 16, Probability and Statistics I.*

(*continued on page 17*)

Solution 3

Solution 3

(i) $E((X-\mu)^2) = \sum_{r=1}^{n} (a_r - \mu)^2 p_r$

$$= \sum_{r=1}^{n} (a_r^2 - 2\mu a_r + \mu^2) p_r$$

$$= \sum_{r=1}^{n} a_r^2 p_r - 2\mu \sum_{r=1}^{n} a_r p_r + \mu^2 \sum_{r=1}^{n} p_r$$

$$= \sum_{r=1}^{n} a_r^2 p_r - 2\mu^2 + \mu^2 \quad \left(\text{as } \sum_{r=1}^{n} p_r = 1 \text{ and } \sum_{r=1}^{n} a_r p_r = \mu\right)$$

$$= E(X^2) - \mu^2$$

(ii) $E((X-\mu)^2) = E(X^2) - \mu^2$

$$= \tfrac{91}{6} - \tfrac{49}{4}$$

$$= \tfrac{35}{12}$$

As expected for a uniform distribution, the variance of X is just the variance of the set $\{1, 2, 3, 4, 5, 6\}$.

(iii) $E(X) = \frac{1}{4} + \frac{6}{4} + \frac{2}{8} + \frac{3}{8} + \frac{4}{8} + \frac{5}{8}$

$$= \tfrac{7}{2} = \mu$$

$$E(X^2) = \tfrac{1}{4} + \tfrac{36}{4} + \tfrac{4}{8} + \tfrac{9}{8} + \tfrac{16}{8} + \tfrac{25}{8}$$

$$= 16$$

so that

$$E(X^2) - \mu^2 = 16 - \tfrac{49}{4}$$

$$= \tfrac{15}{4}$$

(iv) $\mu_Y = E(Y) = E(cX + d)$

$$= \sum_{r=1}^{n} (ca_r + d) p_r \quad \left(\text{for } E(g(X)) = \sum_{r=1}^{n} g(a_r) p_r\right)$$

$$= c \sum_{r=1}^{n} a_r p_r + d \sum_{r=1}^{n} p_r$$

$$= cE(X) + d$$

$$= c\mu_X + d$$

In particular, we notice by putting $d = 0$ that $E(cX) = cE(X)$.

(v) $\operatorname{var}(Y) = E((Y - \mu_Y)^2)$

$$= E((cX + d - c\mu_X - d)^2)$$

$$= E(c^2(X - \mu_X)^2)$$

$$= c^2 \operatorname{var}(X)$$

■

(*continued from page 15*)

For convenience, we denote the outcomes of the N trials by $\{x_1, x_2, \ldots, x_N\}$ and the mean of this set by $\bar{x}$. Then

$$\begin{aligned}\mu = E(X) &= \sum_{r=1}^{n} a_r p_r \\ &\simeq \sum_{r=1}^{n} \frac{a_r m_r}{N} \\ &= \sum_{s=1}^{N} \frac{x_s}{N} \\ &= \bar{x},\end{aligned}$$

the mean of the outcomes of the N trials (the x values) considered simply as data.

$$\begin{aligned}\text{Again, } \operatorname{var}(X) &= \sum_{r=1}^{n} (a_r - \mu)^2 p_r \\ &\simeq \sum_{r=1}^{n} \frac{(a_r - \mu)^2}{N} m_r \\ &\simeq \sum_{r=1}^{n} \frac{m_r(a_r - \bar{x})^2}{N} \qquad \text{(from above)} \\ &= \sum_{s=1}^{N} \frac{(x_s - \bar{x})^2}{N} \\ &= \text{variance of the } x \text{ values considered as data.}\end{aligned}$$

In *Unit 16* we saw that the variance of data is used as a measure of spread of the data; we would therefore expect the variance of a distribution to be a measure of spread of the distribution. "Spread" conveys the idea of "width" or "linear extension", but var (X) is obtained as the sum of quadratic expressions. We therefore use σ $(= \overset{+}{\sqrt{\operatorname{var}(X)}})$ rather than σ^2 $(= \operatorname{var}(X))$ as a measure of spread, and call it the standard deviation of the distribution. Thus

Definition 4
* * *

$$\begin{aligned}\sqrt{E((X - \mu)^2)} &= \sigma \\ &= \text{standard deviation}\end{aligned}$$

In Exercise 3 we found that the fair and the loaded dice had equal means but different variances — and hence different spreads. But we do not just passively observe that different distributions have different spreads; the point is of vital importance in the Theory of Estimation which we discuss in section 21.2.3.

We have now constructed some machinery for coping with simplified models of probabilistic experiments. This machinery will be used later in this unit, as well as in future years.

Comparing Two Distributions

Discussion
* *

In many probability situations we are faced with the problem of choice. Do we go in for this method of selection or that? Do we adopt this trial or that? Do we use this method or that for estimating an unknown value? However we choose, we shall be considering some distribution or other, and very often it comes down to choosing between distributions. Which is the more suitable for our purposes? Even taking the values a_r for granted, it still takes $n - 1$ numbers to specify the distribution (why not n?). Hence we could be left trying to compare two distributions, each having $n - 1$ measures. Somehow we must find a way of representing each distribution by a single measure. This is where expectations come in; for the distributions which we commonly meet have just one mean, so we can decide

between two distributions by comparing their means (if indeed the mean is a relevant measure in the context). If the mean is not relevant, perhaps the variance may be relevant; if so, we have a single measure for comparison purposes.

Example 1

Example 1

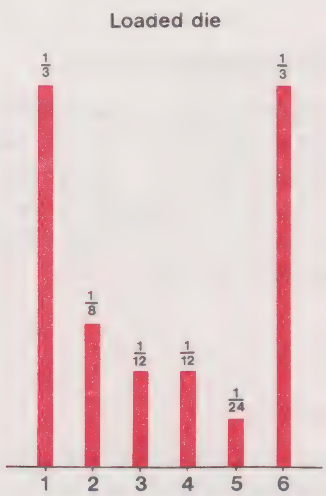
To fix our ideas, suppose we wish to play a simple dice game with an opponent, and we are offered a choice between a fair die having a probability distribution given by

$$P(X = r) = \tfrac{1}{6} \qquad (r = 1, 2, \ldots, 6)$$

and a loaded die having a probability distribution given by

$$P(Y = 1) = P(Y = 6) = \tfrac{1}{3}$$
$$P(Y = 2) = \tfrac{1}{8}$$
$$P(Y = 3) = P(Y = 4) = \tfrac{1}{12}$$
$$P(Y = 5) = \tfrac{1}{24}$$

The two distributions can be illustrated by bar charts:



If the object of the game is to throw the higher score, which of the dice should we choose? That is, which distribution suits us best? We may think on the following lines.

(i) A 6 is a very good number to get. The probability of throwing a 6 with the loaded die is twice the probability of throwing a 6 with the fair die, and therefore it is the better die to have.
However, the probability of throwing a 1 with the loaded die is twice that for the fair die, and this cancels out the advantage mentioned above. It therefore looks as if we shall have to consider the *whole* probability distribution for the loaded die, not just part of it.

(ii) If we do consider the whole distribution for the loaded die, it is surely relevant to consider the mean score. If the expected score from the loaded die is higher than that for the fair die, we should prefer to have it.
Now for the fair die,

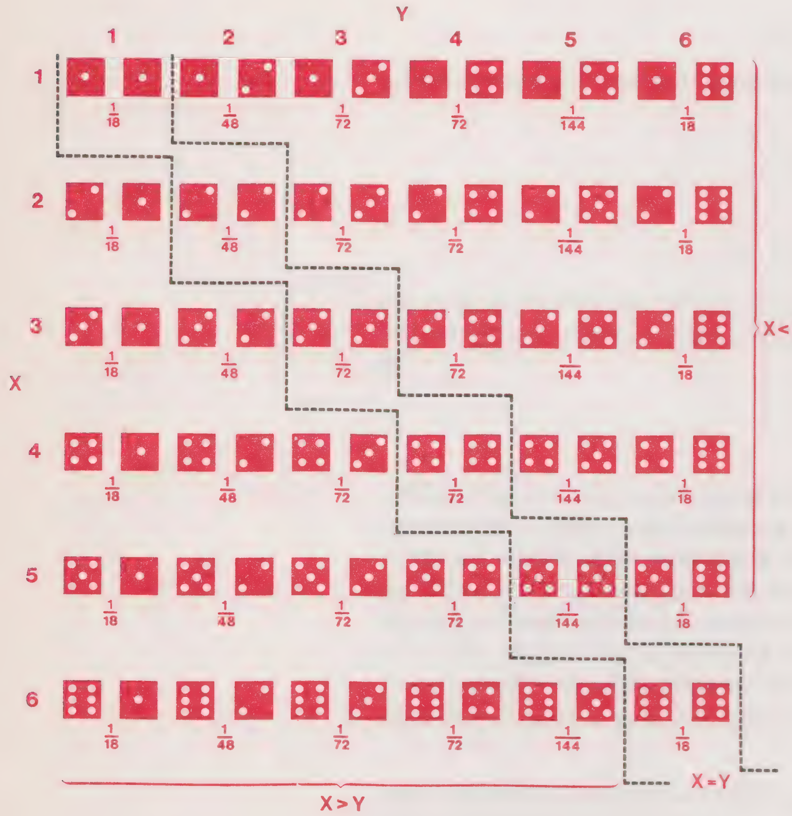
$$E(X) = 3.5$$

For the loaded die,

$$E(Y) = 1 \times \tfrac{1}{3} + 2 \times \tfrac{1}{8} + 3 \times \tfrac{1}{12} + 4 \times \tfrac{1}{12} + 5 \times \tfrac{1}{24} + 6 \times \tfrac{1}{3}$$
$$= \frac{(8+6+6+8+5+48)}{24}$$
$$= \tfrac{81}{24} = 3.375 < E(X).$$

Therefore it is better to have the fair die.

(iii) Argument (ii) seems relevant, but does it settle the matter beyond all further consideration? The only way to find out is to examine all possible pairs of outcomes, in which the first and second elements of the pair represent the outcomes of throwing the fair die and the loaded die respectively. These are shown below with the probabilities attached.



Because the events "$X = r$" and "$Y = s$" are statistically independent, the probabilities in this table have been obtained by multiplication. For example,

$$P(X = 1; Y = 1) = P(X = 1) \times P(Y = 1) = \tfrac{1}{6} \times \tfrac{1}{3} = \tfrac{1}{18}$$

By adding up the probabilities in the three regions of the table, we find that

$$P(X > Y) = \tfrac{21}{48},$$
$$P(X = Y) = \tfrac{8}{48},$$
$$P(X < Y) = \tfrac{19}{48},$$

Equations (1)

where, for instance, $P(X > Y)$ is the probability that the realization of X is greater than the realization of Y.

The sum of these probabilities is 1, reflecting the fact that the three events are mutually exclusive and exhaust all the possible outcomes of the trial.

Equations (1) show that the fair die is "better" than the loaded die in the sense that, if they are both thrown once, the fair die is more likely to yield the larger score than the smaller.

This backs up the conclusion reached in (ii), and this might have been expected; but it is not all that "obvious". The higher mean does *not necessarily* indicate that we are more likely to get the higher score with the fair die (see the following exercise). ■

Exercise 4

Exercise 4
(3 minutes)

Let X denote the score obtained from a loaded die, where the probability distribution of X is given by:

$$P(X = r) = \tfrac{1}{4} \qquad (r = 1, 2, 3)$$

and

$$P(X = r) = \tfrac{1}{12} \qquad (r = 4, 5, 6).$$

Let Y denote the score obtained from another loaded die with probability distribution given by:

$$P(Y = r) = \tfrac{35}{72} \qquad (r = 2, 3)$$

and

$$P(Y = r) = \tfrac{1}{144} \qquad (r = 1, 4, 5, 6).$$

Show that, if each die is thrown once, the first die is less likely to show the larger score than the second, but that the expectation of score for the first die is larger than for the second. ■

Summary

Summary
*

We have learnt from this section that it is not always immediately possible to choose between two distributions, because there may be more than one measure relevant to the situation. But if it is quite clear what is meant by "better", it may be possible to find just one characteristic of a distribution which can be used as a criterion for the choice. In our Example 1 this single measure was the distribution mean. In other circumstances, the more relevant single measure could be the distribution variance. Although these two measures do not exhaust the possibilities, they are the two in most common use.

However, a consideration of the mean could lead us astray (as we have seen in Exercise 4); it is therefore safer to regard any particular single standard measure as *indicating* the correct conclusion rather than settling it beyond all doubt. It is always dangerous in statistics to apply standard techniques blindly.

21.2 SAMPLING AND ITS CONSEQUENCES

21.2

21.2.1 Random Sampling

21.2.1

Discussion
* *

If we throw a die which is known to be fair, then we have a probability model of the situation, and we can make predictions about outcomes. On the other hand, we may be faced with the very question: "Is the die fair?" A knowledge of the sort of results which would result from throwing both fair and loaded dice will clearly be a help in answering this question. Just how it will help is a matter for statistical theory.

To investigate this question, we would have to experiment with the die, i.e., throw it a number of times. The possible outcomes of any single trial are the numbers 1, 2, 3, 4, 5, 6. Let

x_1 be the outcome of the first trial;

x_2 be the outcome of the second trial;

$\vdots$

x_n be the outcome of the nth trial.

x_1 is the numerical value taken by the random variable X in the first trial; it is the realization of X on this occasion. Similarly for x_2, and so on. The sequence of values $x_1, x_2, \ldots, x_n$ is then called a random sample from the probability distribution of X. It is important to observe two properties of random samples:

Definition 1
* * *

(i) each sample value x_r is a realization of the same random variable X;
(ii) the realizations for the various trials are physically independent; therefore they are statistically independent (see *Unit 18*).

Thus $P(X = x_1 \text{ in first trial}; X = x_2 \text{ in second})$
$= P(X = x_1 \text{ in first trial}) \times P(X = x_2 \text{ in second})$
$= P(X = x_1) \times P(X = x_2)$,
since X has the same probability distribution in all trials. For the whole sample, the probability of getting the sequence $x_1, x_2, \ldots, x_n$ is

$$P(X = x_1) \times P(X = x_2) \times \cdots \times P(X = x_n).$$

Remarks (i) and (ii) above are applicable to *any* random variable defined on *any* finite sample space.

The concept of a random sample has so far been expressed in terms of a sequence of identical and independent trials; for example, a die thrown n times. The same model would apply equally to a set of simultaneous identical and independent trials; for example, n distinguishable dice all thrown at once. It is often convenient to regard a random sample $x_1, x_2, \ldots, x_n$ as if each observation x_i were a realization of a distinct random variable X_i. We then get n random variables all with the same probability distribution. These two ways of regarding a random sample are equivalent. The second is usually the more useful mathematically.

Randomness, like independence, has been expressed purely in mathematical terms. To get observations which may reasonably be supposed to conform to it, precautions often need to be taken. These may be obvious to a good experimenter; for example, a chemist making a series of chemical determinations would usually wash his test-tubes so that there was no risk of contamination between experiments. On the other hand, such precautions may need to be more subtle or may even be impossible. For example, in three successive landings of an aircraft it is impossible to believe that the chance of failure of a tyre is *exactly* the same on each occasion (we cannot disregard tyre wear or even pilot fatigue). It may, however, be possible, by tyre and/or pilot changes, to arrange the experiment so that the three observations obtained may reasonably be supposed to be a random sample from the population which is of interest.

(*continued on page 22*)

Solution 21.1.4.4

Solution 21.1.4.4

For X and Y alone the probability distributions are given by:

	1	2	3	4	5	6
X	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
Y	$\frac{1}{144}$	$\frac{35}{72}$	$\frac{35}{72}$	$\frac{1}{144}$	$\frac{1}{144}$	$\frac{1}{144}$

$$\left.\begin{aligned} E(X) &= \tfrac{11}{4} = 2.75 \\ E(Y) &= \tfrac{183}{72} = 2.54 \end{aligned}\right\} \text{so } E(X) > E(Y).$$

The distribution of X and Y is obtained as in the text:

X \ Y	1	2	3	4	5	6
1	$\frac{1}{576}$	$\frac{35}{288}$	$\frac{35}{288}$	$\frac{1}{576}$	$\frac{1}{576}$	$\frac{1}{576}$
2	$\frac{1}{576}$	$\frac{35}{288}$	$\frac{35}{288}$	$\frac{1}{576}$	$\frac{1}{576}$	$\frac{1}{576}$
3	$\frac{1}{576}$	$\frac{35}{288}$	$\frac{35}{288}$	$\frac{1}{576}$	$\frac{1}{576}$	$\frac{1}{576}$
4	$\frac{1}{1728}$	$\frac{35}{864}$	$\frac{35}{864}$	$\frac{1}{1728}$	$\frac{1}{1728}$	$\frac{1}{1728}$
5	$\frac{1}{1728}$	$\frac{35}{864}$	$\frac{35}{864}$	$\frac{1}{1728}$	$\frac{1}{1728}$	$\frac{1}{1728}$
6	$\frac{1}{1728}$	$\frac{35}{864}$	$\frac{35}{864}$	$\frac{1}{1728}$	$\frac{1}{1728}$	$\frac{1}{1728}$

$X < Y$

$X > Y$

$X = Y$

From the table, we find that

$$P(X > Y) = \tfrac{107}{288}$$

$$P(X < Y) = \tfrac{110}{288}$$

and so we see that $P(X > Y) < P(X < Y)$, although $E(X) > E(Y)$. ■

(*continued from page 21*)

Exercise 1

Exercise 1
(3 minutes)

(i) A fair die is thrown twice. Write down all the possible random samples, together with their probabilities. (Each random sample is, of course, a sequence of two elements.)
(ii) Two indistinguishable fair dice are thrown simultaneously. What difference, if any, would this make to the results obtained above?
(iii) An experimenter throws two fair dice, and records the larger score (or the common score if the two scores are equal). He throws them a second time, and again records the larger score. Thus he finishes with a sequence of two recorded scores. Write down all the possible random samples together with their probabilities. ■

Discussion
* *

Exercise 1 is so similar to Example 21.1.2.1 on page 7 that you may be wondering if the definition of a random sample really introduces anything new. To help you understand the importance of the concept of a random sample, we shall now consider how to construct random samples from known probability distributions using random numbers.

Suppose that we know the sample space of an experiment, but we do not know what the sequence of results would be if we were to perform the experiment several times. If we want to simulate the experiment, we need a method of producing such a sequence in a way which reflects the randomness of the experiment and the probabilities assigned to each of the possible outcomes of the experiment.

Suppose that we have a random variable X which can take values $a_1, a_2, \ldots, a_k$ and that the probability distribution of X is given by

$$P(X = a_r) = \frac{b_r}{c} \qquad (r = 1, 2, \ldots, k)$$

where $b_1, b_2, \ldots, b_k$ are positive integers, and c is their sum.

A table of random numbers consists of a sequence of digits obtained in a way equivalent to throwing successively a ten-sided fair die whose faces are numbered 0, 1, 2, ..., 9. Suppose that $c \leqslant 10$. Then to obtain a random sample, $x_1, x_2, \ldots, x_n$, from the known probability distribution, we read n successive digits from the random numbers. If n_r is the rth such random number, then

$$\begin{aligned} &\text{if } 0 \leqslant n_r \leqslant b_1 - 1 && \text{we take } x_r = a_1; \\ &\text{if } b_1 \leqslant n_r \leqslant b_1 + b_2 - 1 && \text{we take } x_r = a_2; \\ &\text{if } b_1 + b_2 \leqslant n_r \leqslant b_1 + b_2 + b_3 - 1 && \text{we take } x_r = a_3; \end{aligned}$$

and so on. If any number obtained from the table is greater than or equal to c, we ignore it and continue with the next random number.

Example 1

Example 1

The following is a table of 100 random digits:

4	6	2	0	5	7	6	7	9	9
1	3	1	7	5	6	4	4	5	1
3	3	2	3	3	5	7	3	8	6
5	1	9	5	5	3	0	5	5	7
6	7	7	0	7	4	5	8	7	8
2	8	8	7	4	9	1	0	2	6
1	9	1	8	1	8	4	6	0	2
4	4	1	9	2	2	9	3	8	6
6	1	2	2	7	0	4	0	1	9
6	2	3	1	8	7	4	0	1	1

Let X be the random variable corresponding to throwing a fair die, so that

$$a_r = r \qquad (r = 1, 2, \ldots, 6)$$

$$P(X = a_r) = \tfrac{1}{6}$$

In the notation we introduced above, we have

$$b_r = 1 \qquad (r = 1, 2, \ldots, 6)$$

$$c = 6$$

We now obtain a random sample using the above rules. The first random number in the table is $n_1 = 4$, and

$$4 = b_1 + b_2 + b_3 + b_4 \leqslant n_1 \leqslant b_1 + b_2 + b_3 + b_4 + b_5 - 1,$$

so that x_1, the first element in our sequence, is 5. The second random number is 6 and, following our instructions, we ignore it. In general, if $n_r < c$, we find that $x_r = n_r + 1$, and we get the random sample

5, 3, 1, 6, 2, 4, 2, ... ■

If $10 < c \leqslant 100$, we have to obtain random two-digit numbers from the table; we do this by reading the random digits in pairs. Apart from this modification, the above instructions hold.

(continued on page 25)

Solution 1

Solution 1

(i) This solution has in effect already been given at the beginning of section 21.1.2 (page 7).

(ii) The only difference here is that there is now no distinction between the dice; that is, a random sample is a sequence having only one element, that element being a pair of numbers (not an *ordered* pair); we therefore write the element which consists of the numbers a and b as the *set* $\{a, b\}$.*

$\{1,1\}$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{1,5\}$	$\{1,6\}$
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
	$\{2,2\}$	$\{2,3\}$	$\{2,4\}$	$\{2,5\}$	$\{2,6\}$
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
		$\{3,3\}$	$\{3,4\}$	$\{3,5\}$	$\{3,6\}$
		$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
			$\{4,4\}$	$\{4,5\}$	$\{4,6\}$
			$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
				$\{5,5\}$	$\{5,6\}$
				$\frac{1}{36}$	$\frac{2}{36}$
					$\{6,6\}$
					$\frac{1}{36}$

(iii) In order to see the general pattern, it is helpful to consider a few special cases:

first throw		second throw		random sequence
die A	die B	die A	die B	
1	1	1	1	1, 1
1	1	1	2	1, 2
1	1	2	1	1, 2
1	1	2	2	1, 2

The random sequence 1, 1 can only be obtained in 1 way; the random sequence 1, 2 can be obtained in 3 ways; and so on. Since each experiment involves 4 throws, the sequence 1, 1 has probability $\frac{1}{6^4}$; the sequence 1, 2 has probability $\frac{3}{6^4}$; and so on.

The samples with their probabilities are:

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
$\frac{1}{1296}$	$\frac{3}{1296}$	$\frac{5}{1296}$	$\frac{7}{1296}$	$\frac{9}{1296}$	$\frac{11}{1296}$
2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
$\frac{3}{1296}$	$\frac{9}{1296}$	$\frac{15}{1296}$	$\frac{21}{1296}$	$\frac{27}{1296}$	$\frac{33}{1296}$
3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
$\frac{5}{1296}$	$\frac{15}{1296}$	$\frac{25}{1296}$	$\frac{35}{1296}$	$\frac{45}{1296}$	$\frac{55}{1296}$
4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
$\frac{7}{1296}$	$\frac{21}{1296}$	$\frac{35}{1296}$	$\frac{49}{1296}$	$\frac{63}{1296}$	$\frac{77}{1296}$
5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
$\frac{9}{1296}$	$\frac{27}{1296}$	$\frac{45}{1296}$	$\frac{63}{1296}$	$\frac{81}{1296}$	$\frac{99}{1296}$
6, 1	6, 2	6, 3	6, 4	6, 5	6, 6
$\frac{11}{1296}$	$\frac{33}{1296}$	$\frac{55}{1296}$	$\frac{77}{1296}$	$\frac{99}{1296}$	$\frac{121}{1296}$

■

* In *Unit 1, Functions* we noted that, when listing the elements of a set, we do not record repetitions. "Repetitions" in that context really means "unnecessary repetitions"; in this context, we write the element consisting of the pair of numbers a and a as $\{a, a\}$.

(*continued from page 23*)

The possibility of short cuts should not be overlooked. If we use the method described above to construct random samples from the distribution given by

$$P(X = 0) = \tfrac{1}{5}$$

$$P(X = 1) = \tfrac{4}{5},$$

we have $c = 5$, and half the numbers in the random table are therefore wasted. It is better in this case to re-express the probabilities as

$$P(X = 0) = \tfrac{2}{10}$$

$$P(X = 1) = \tfrac{8}{10}$$

We now have $c = 10$; the process is less wasteful than before and easier to apply.

A more sophisticated version of this technique, suitable for use on computers, is used under the name *simulation* to examine the consequences of probability models in situations where the calculations are very complicated; it also has a minor place in the history of statistical theory where some of the standard results, now established mathematically, were first obtained by this procedure.

The next exercise is intended to show that a small sample may not be adequate to distinguish between two probability models.

Exercise 2

Exercise 2
(2 minutes)

Use the random numbers given in Example 1 to obtain:

(i) a random sample consisting of 20 scores obtained from a fair die (see Example 1);

(ii) a random sample consisting of 20 scores obtained from a loaded die, whose probability distribution is given by

$$P(X = 1) = P(X = 6) = \tfrac{1}{5}$$

$$P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = \tfrac{3}{20}$$

In this second case, begin reading the random numbers from the beginning of the next full row from where you left off in (i), and take $c = 100$. ■

Solution 2 **Solution 2**

(i) The random sample is

$$5, 3, 1, 6, 2, 4, 2, 6, 5, 5, 6, 2, 4, 4, 3, 4, 4, 6, 4, 6.$$

This has used the first 31 digits.

(ii) (Notice that with $c = 20$, we wouldn't have enough digits left in the table.)

$c = 100$, so we use two digits at a time from the table.

We have

$$b_1 = 20, b_2 = b_3 = b_4 = b_5 = 15, b_6 = 20$$

and

$$a_r = r \qquad (r = 1, 2, \ldots, 6).$$

Since $b_1 = 20$, for any number n_r from the table less than or equal to 19, we have

$$0 \leqslant n_r \leqslant b_1 - 1 = 19, \text{and so } x_r = a_1 = 1.$$

Proceeding in a similar way, we can obtain the table:

	score
$0 \leqslant n_r \leqslant 19$	1
$20 \leqslant n_r \leqslant 34$	2
$35 \leqslant n_r \leqslant 49$	3
$50 \leqslant n_r \leqslant 64$	4
$65 \leqslant n_r \leqslant 79$	5
$80 \leqslant n_r \leqslant 99$	6

We now use the digits in pairs beginning at the fifth row (i.e. we use 67, 70, 74, . . .) and the random sample is

$$5, 5, 5, 4, 5, 2, 6, 3, 1, 2, 1, 1, 1, 3, 1, 3, 1, 2, 6, 6.$$

The following table shows the frequency with which each score occurs in the random sample.

Score	1	2	3	4	5	6
(i) Fair die	1	3	2	6	3	5
(ii) Loaded die	6	3	3	1	4	3

If anything, the results for the loaded die look more evenly distributed than those for the fair die. These numbers were *not* specially chosen to exhibit this property. ■

21.2.2 Sampling Statistics and Their Distributions

21.2.2

Introduction

Introduction
* *

Given a random sample, $x_1, x_2, \ldots, x_n$, from some sample space, it will have a mean $\bar{x}$. If we repeated the whole experiment, we would have another random sample $x'_1, x'_2, \ldots, x'_n$ with mean $\bar{x}'$. We could repeat the whole operation many times, getting further means $\bar{x}''$, $\bar{x}'''$ and so forth. Now we obtained the sample $x_1, x_2, \ldots, x_n$ by carrying out a trial n times, the rth trial providing the value x_r. If we think of these n trials as constituting

one single "super-trial", $\bar{x}$ is the outcome of this super-trial. Other super-trials give us the outcomes $\bar{x}', \bar{x}'', \bar{x}'''$; but this situation is no different in kind from ordinary trials; therefore $\bar{x}, \bar{x}', \bar{x}'', \ldots$ are realizations of a random variable which may be denoted by $\bar{X}$. $\bar{X}$ is a special case of a *sampling statistic*, and the probability distribution of $\bar{X}$ is referred to as a *sampling distribution*. This nomenclature makes reference to the sampling process which gives rise to $\bar{X}$, but it also has the advantage of distinguishing $\bar{X}$ from the original random variable X, and the sampling distribution of $\bar{X}$ from the original (often called *parent*) distribution of X. In other words, the "children" of *parent distributions* are *sampling distributions*.

Sampling Statistics

In the notation introduced in the previous paragraph, we can define a function which maps a random sample, $x_1, x_2, \ldots, x_n$, to its mean $\bar{x}$. More generally, if there is a function

$$h:(x_1, x_2, \ldots, x_n) \longmapsto h((x_1, x_2, \ldots, x_n)) = y$$

from the set of all random samples to R, then in the same way as for the particular case $\bar{x}$, y is a realization of a random variable Y which is called a sampling statistic. The probability distribution of Y is called a sampling distribution. The subject of mathematical statistics is mainly concerned with the determination of these sampling distributions.

Definition 1

Definition 2

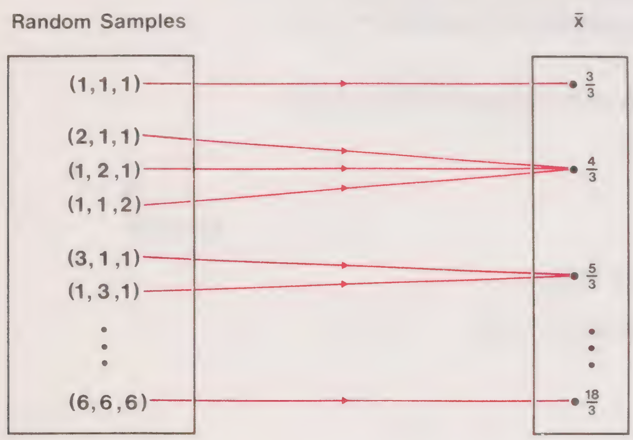
Example 1

Example 1

Returning to the special case of the mean, suppose we throw a fair die three times, and that we are interested in the arithmetic mean of the values x_1, x_2, x_3 thus obtained. In other words, we are interested in

$$y = \frac{x_1 + x_2 + x_3}{3} = \bar{x}$$

and in the sampling distribution of $\bar{X}$. There are 6 possible values for each of x_1, x_2, x_3, and therefore there are $6^3 = 216$ possible samples. Only the sample 1, 1, 1 has 1 as its mean. The samples 2, 1, 1; 1, 2, 1; 1, 1, 2 all have means of $\frac{4}{3}$. The mapping from samples to their means is shown below:



By evaluating the means of all the 216 sample points and remembering that each occurs with a probability of $\frac{1}{216}$, we find that the sampling statistic $\bar{X}$ has the probability distribution:

$\bar{x}$	$\frac{3}{3}$	$\frac{4}{3}$	$\frac{5}{3}$	$\frac{6}{3}$	$\frac{7}{3}$	$\frac{8}{3}$	$\frac{9}{3}$	$\frac{10}{3}$	$\frac{11}{3}$	$\frac{12}{3}$	$\frac{13}{3}$	$\frac{14}{3}$	$\frac{15}{3}$	$\frac{16}{3}$	$\frac{17}{3}$	$\frac{18}{3}$
$P(\bar{X} = \bar{x})$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

■

Exercise 1

Exercise 1
(5 minutes)

A fair die is thrown k times, and x_{max} is the greatest of the k scores thus obtained. If X_{max} is the random variable having x_{max} as its realization, show that X_{max} has distribution given by

$$P(X_{max} = r) = \left(\frac{r}{6}\right)^k - \left(\frac{r-1}{6}\right)^k, \qquad (r = 1, 2, \ldots, 6).$$

(HINT: Work out the answer first for $x_{max} = 1$, then for $x_{max} = 2$, etc.)

■

Discussion
* *

Sampling statistics are simply random variables, but they are based on random samples rather than on outcomes of individual trials; other than this, they do not raise any essentially new ideas. If, as in the simple cases which we have considered, the complete sampling distribution can be determined, then we have complete information for making probability statements about the sampling statistic. In complicated cases, we cannot obtain the sampling distribution, and we have to be content with knowing only its mean and variance; these measures are sometimes referred to as the sample mean and the sample variance of the statistic.

The measures for the parent and sampling distribution are not unconnected. Let us look at the mean and variance of the sampling distribution of the average score for three throws of a fair die (Example 1). Using that distribution, we obtain

$$E(\bar{X}) = \sum_{r=3}^{18} \frac{r}{3} P\left(\bar{X} = \frac{r}{3}\right) = \frac{2268}{648} = 3.5$$

Thus the mean of that sampling distribution is the same (3.5) as the mean which we obtained on page 14 for the parent distribution.

At this stage, it is as well to take stock so as to avoid verbal confusion. Given any random variable X, it has a mean $\mu = E(X)$. Given a random sample, the sampling statistic $\bar{X}$ (whose realizations are sample means $\bar{x}$) has a probability distribution whose mean is $E(\bar{X})$. Thus in a sense we have the "mean of the mean". There is no reason why any particular realization $\bar{x}$ should equal μ; and we have seen no reason so far for connecting $E(\bar{X})$ with μ. For the die, however, we have demonstrated for samples of three that $E(\bar{X}) = 3.5 = \mu$.

Similarly, we can calculate the mean of the sample variance, or the variance of the sample mean.

Example 2

Example 2

For three throws of a fair die, we shall find var $(\bar{X})$.

$$\begin{aligned}
\text{var}(\bar{X}) &= E(\bar{X}^2) - \mu^2 \qquad \text{(see Exercise 21.1.4.3)} \\
&= \sum_{r=3}^{18} \frac{r^2}{9} P\left(\bar{X} = \frac{r}{3}\right) - \frac{49}{4} \\
&= \frac{9}{9} \times \frac{1}{216} + \frac{16}{9} \times \frac{3}{216} + \frac{25}{9} \times \frac{6}{216} + \cdots \\
&\quad + \frac{324}{9} \times \frac{1}{216} - \frac{49}{4} \\
&= \frac{25704}{1944} - \frac{49}{4} \\
&= \frac{35}{36}
\end{aligned}$$

■

This result may be compared with the result for the original probability distribution of X, for the score of a fair die, where we obtained var $(X) = \frac{35}{12}$. (See Exercise 21.1.4.3(ii).) Thus, by taking the average of a sample of three observations, the variance has been reduced by a factor of 3. This is not just a coincidence; it is a special case of a general rule about the variance of the sampling distribution of the mean. We shall see the practical importance of this in section 21.2.3, where we look at estimation.

The computations of the mean and variance of $\bar{X}$ in this simple case suggest certain general properties. These general properties will not be proved in this unit, but we state them here.

If $E(X) = \mu$ and var $(X) = \sigma^2$,

then for a random sample of size n,

$$E(\bar{X}) = \mu \text{ and var } (\bar{X}) = \frac{\sigma^2}{n}$$

Since variance is used as a measure of dispersion, these results imply that the realizations of $\bar{X}$ are more closely clustered — and hence closer to μ — the larger n becomes.

Some of the material in this and earlier sections is summarized below.

Summary
* *

(i) A random variable X has realizations $a_1, a_2, \ldots, a_k$ which are possible numerical outcomes of a single trial.

(ii) A sampling statistic is a random variable associated with a random sample; it is some function h defined on the set of random samples.

	Equal to	Denoted by
(iii) expectation of X = mean of distribution of X	$\sum_{r=1}^{k} a_r P(X = a_r)$	$E(X)$ or μ
(iv) mean of $x_1, x_2, \ldots, x_n$	$\sum_{r=1}^{n} \frac{x_r}{n}$	$\bar{x}$
(v) mean of random variables $X_1, \ldots, X_n$	$\sum_{r=1}^{n} \frac{X_r}{n}$	$\bar{X}$
(vi) variance of $x_1, x_2, \ldots, x_n$	$\sum_{r=1}^{n} \frac{(x_r - \bar{x})^2}{n}$	s^2
(vii) variance of X	$E(X - \mu)^2$	σ^2 or var (X)
(viii) mean of sample means	μ*	$E(\bar{X})$
(ix) variance of sample means	$\frac{\sigma^2}{n}$*	var $(\bar{X})$

* These two results have been quoted without proof.

Solution 1

Solution 1

$$P(X_{\max} = 1) = P(\text{all } k \text{ scores are } 1)$$
$$= \left(\frac{1}{6}\right)^k$$
$$= \left(\frac{1}{6}\right)^k - \left(\frac{0}{6}\right)^k$$

$$P(X_{\max} = 1) + P(X_{\max} = 2) = P(X_{\max} \leqslant 2)$$
$$\therefore \quad P(X_{\max} = 2) = P(X_{\max} \leqslant 2) - P(X_{\max} = 1)$$
$$= P(\text{all } k \text{ scores are} \leqslant 2) - \left(\frac{1}{6}\right)^k$$
$$= \left(\frac{2}{6}\right)^k - \left(\frac{1}{6}\right)^k$$

Similarly

$$P(X_{\max} = 3) = P(X_{\max} \leqslant 3) - P(X_{\max} \leqslant 2)$$
$$= \left(\frac{3}{6}\right)^k - \left(\frac{2}{6}\right)^k$$

In general,

$$P(X_{\max} = r) = P(X_{\max} \leqslant r) - P(X_{\max} \leqslant r - 1)$$
$$= \left(\frac{r}{6}\right)^k - \left(\frac{r-1}{6}\right)^k$$

■

21.2.3 Parametric Models and Statistics as Estimates

21.2.3

Discussion

So far we have assumed that we knew the probabilities of the various outcomes in the sample space corresponding to a single trial. We have, for example, assumed that a die being thrown is fair, or if it is loaded, that the form of the bias is known; that the probability of failure of an aircraft tyre is known; and so on. In real life the nature of the bias in a die or the probability of failure of a tyre will be unknown. It is precisely these unknowns about which we hope to obtain information from our observations. To do this we need first to construct models in which a statement such as "the die is loaded" is made more specific. This is usually done by the introduction of one or more *parameters.*

If we know nothing about a die, all we can say about the random variable corresponding to the score is that it satisfies

$$P(X = r) = p_r \qquad (r = 1, 2, \ldots, 6)$$

where $0 \leqslant p_r \leqslant 1$ and $\sum_{r=1}^{6} p_r = 1$;

Definition 1

$p_1, p_2, \ldots, p_6$ are called parameters of the system. They are constant in any situation (e.g. for the same die thrown in the same kind of way), though their values are unknown. The distribution defined above is our probability model of the situation. Note that models are not necessarily meant to be absolutely correct in every detail. Given points on a graph, a scientist does not necessarily find a curve to pass accurately through each and every point. If the data appear to warrant it, he fits a straight line which represents the essential trend indicated by the points. In our case, we try to use any information about the physical system to *reduce* the number of parameters.

For instance, suppose that we have a die for which there is some evidence that it is biassed, because the 6's do not occur one in six times in a long sequence of trials although all the other scores appear to occur equally often. If we take the probability of a 6 as p, and the probabilities of all the other numbers to be equal (or even approximately equal), then the appropriate model to take is the distribution given by

$$P(X = 6) = p$$

$$P(X = r) = \frac{1-p}{5} \qquad (r = 1, 2, \ldots, 5)$$

On the other hand, we may know something of the construction of the die, and we may think that the following model is more accurate:

$$P(X = 6) = p$$

$$P(X = r) = \tfrac{1}{6} \qquad (r = 2, 3, 4, 5)$$

$$P(X = 1) = \tfrac{1}{3} - p$$

where $0 \leqslant p \leqslant \frac{1}{3}$ but p is otherwise unknown.

Both these models contain a single parameter, because we have taken into account our knowledge of the system. (In much the same way, a scientist may fit a straight line to a set of points because other reasoning leads him to expect a straight line, and the object of his experiment is to find a particular line rather than *any* curve to fit the data.)

In general, the problem of choosing an appropriate model requires a great deal of background knowledge about the physical conditions of the experiment, as well as experience of analogous physical situations. The use of standard statistical methods often obscures the need to examine the appropriateness of the model.

The determination of the sampling distribution of a statistic based on observations from a parametric model again presents no new problems in principle. What is new in the situation is that there is little point in taking the statistic *unless* it provides some information about the value of the parameter. From any particular sample it is possible to compute many different sampling statistics, and the nature of the information which any particular statistic provides about the parameter can only be determined from its sampling distribution. The first role sampling distributions play in statistical theory is therefore in the comparison of sampling statistics as estimates of the parameter (or parameters) in the model.

Example 1 **Example 1**

As an example to illustrate this, consider a situation which at first sight might appear to be modelled by an N-sided fair die for which N is unknown. Suppose an enemy produces pieces of equipment, and each piece of equipment receives a serial number. Assuming that these serial numbers run from 1 to N, intelligence officers obtain the serial numbers of n of these pieces of equipment, and require information about N, the total number of pieces of equipment made. Alternatively, suppose a forger has produced banknotes numbered from 1 to N. After he has been caught, the Bank of England wish to know how many forged notes are in circulation. Again, the information about N will have to be obtained from the serial numbers of n notes which have been detected as forgeries.

It is reasonable to assume that in each of these situations all existing serial numbers are equally likely to be obtained. (This assumption may be false, but here again we face the eternal problem for the statistician. What other assumption could he make?) Note that we assume that no two items have the same serial number; this is not exactly the situation described by a die, since for the die the same score can come up twice. If, as a model, we regard

all the serial numbers as being written on cards which are then shuffled, and the cards drawn one at a time, then the situation we are dealing with is *random sampling without replacement*, whereas throwing a die corresponds to *random sampling with replacement*. (Random selection with and without replacement was discussed in *Unit 18*.) In each of these situations, however, N is a parameter; it is an unknown, whose value needs to be specified if exact probability predictions of the result of an experiment are to be made. In the statistical context it is something which we wish to *estimate* from the data we obtain. The way a statistician does this is to take the realization of some sampling statistic whose distribution is in some sense grouped closely round the true value of N. A sampling statistic with this kind of property is sometimes said to be an estimator of the parameter.

Definition 2 * *

To illustrate the sort of considerations which would enter into our choice of such a statistic, suppose we obtain two serial numbers, x_1 and x_2. This pair of numbers is supposed to be a random sample. As the first obvious attempt to estimate N, consider the average $\bar{x} = \frac{x_1 + x_2}{2}$; this is a realization of the sampling statistic $\bar{X}$. The possible values this statistic can take are $\frac{3}{2}, \frac{4}{2}, \frac{5}{2}, \ldots, \frac{2N-1}{2}$. The sampling distribution can be obtained by relatively simple arguments, but the work is tedious. (For example, to find $E(\bar{X})$ we would have to find all the possible values of $\bar{X}$ and then the associated probabilities. The probabilities can be calculated from the fact that there are $N(N-1)$ possible pairs, each with probability $\frac{1}{N(N-1)}$).
We shall simply quote the results we want, namely:

$$E(\bar{X}) = \frac{N+1}{2}$$

$$E(\bar{X}^2) = \frac{7N^2 + 11N + 4}{24}$$

so that

$$\begin{aligned}\operatorname{var}(\bar{X}) &= \frac{7N^2 + 11N + 4}{24} - \frac{(N+1)^2}{4} \\ &= \frac{N^2 - N - 2}{24}\end{aligned}$$

Now the serial numbers have a distribution whose random variable is X, and the mean of X is $\frac{N+1}{2}$. Further, we have the sampling distribution of $\bar{X}$, and this also has mean $\frac{N+1}{2}$. Then we can either think of our numerical $\bar{x}$ as being the mean of a sample* of 2 from the original distribution, or as being a sample of 1 from the sampling distribution of $\bar{X}$; the two amount to exactly the same in the end, for in either case we are mentally "equating" $\bar{x}$ with $\frac{N+1}{2}$. But if we are prepared to take $\bar{x}$ as our estimate of $\frac{N+1}{2}$, we must be prepared to take $2\bar{x}$ as our estimate of $N+1$, and hence $2\bar{x} - 1$ as our estimate of N. The random variable having $2\bar{x} - 1$ as its realization is $2\bar{X} - 1$; denoting this by $\hat{N}$, we are using the sampling statistic $\hat{N} = 2\bar{X} - 1$ to estimate N.

At this stage, let us take stock of the situation. Out of the $N(N-1)$ possible serial number pairs we have obtained one, and we have taken the mean, $\bar{x}$, of the two numbers constituting the pair. If we repeated the whole

* Remember that the sampling here is sampling with replacement.

thing from scratch (remember that we are sampling without replacement), we would get a mean $\bar{x}_2$. On the nth occasion we would get $\bar{x}_n$. These n values, $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, are realizations of the sampling statistic $\bar{X}$. We have taken $\hat{N} = 2\bar{X} - 1$, and we know that its expected value is equal to N. Although no realization of $\hat{N}$ is necessarily equal to N, it is in a sense centred on N as N is the expectation of $\hat{N}$.

We observe further that X is distributed symmetrically about its mean value of $\frac{N+1}{2}$, so that $\bar{X}$ is distributed symmetrically about its mean value of $\frac{N+1}{2}$. So $\hat{N} = 2\bar{X} - 1$ is distributed symmetrically about its mean value of N. Thus $\hat{N}$ is just as likely to be too small by a certain amount as it is to be too big by that amount. This would appear to be a desirable property of an estimator.

We have argued that

$\hat{N}$ has mean value N;

$\hat{N}$ is distributed symmetrically about N.

We also know that var $(\hat{N})$

$$= \text{var}\,(2\bar{X} - 1) = 4\,\text{var}\,(\bar{X}) \qquad \text{(see Exercise 21.1.4.3.)}$$

$$= \frac{N^2 - N - 2}{6} \qquad \text{(see page 32)}$$

which gives us a measure of the spread of $\hat{N}$.

Another statistic whose sampling distribution may easily be derived is $X^* = \max\,(X_1, X_2)$. From the array of possible samples we may obtain the sampling distribution of X^*, and hence work out the mean and variance of X^*; these are given by

$$E(X^*) = \frac{2(N+1)}{3}$$

$$\text{var}\,(X^*) = \frac{N^2 - N - 2}{18}$$

For the same kinds of reason as before, we would be prepared to accept x^* (a realization of X^*) as an estimate of $E(X^*) = \frac{2(N+1)}{3}$. It follows that we would accept $\frac{3}{2}x^* - 1$ as an estimate of N. This estimate is a realization of the random variable $N^* = \frac{3}{2}X^* - 1$, which has mean

$$E(N^*) = N$$

and variance

$$\begin{aligned} \text{var}\,(N^*) &= \text{var}\,(\tfrac{3}{2}X^* - 1) \\ &= \tfrac{9}{4}\,\text{var}\,(X^*) \\ &= \frac{(N^2 - N - 2)}{8} \end{aligned}$$

So now we have two different estimates of N, both of them shown to be reasonable. Is there anything to choose between them? Both have mean N, so there is nothing to choose between them on that score. However,

$$\text{var}\,(\hat{N}) = \frac{N^2 - N - 2}{6}$$

and

$$\text{var}\,(N^*) = \frac{N^2 - N - 2}{8}$$

Now variance is a measure of dispersion. Hence the smaller the variance, the more compact the distribution, and the nearer the values within the distribution will be to the mean. We know that $\text{var}(N^*) = \frac{3}{4}\,\text{var}(\hat{N})$, and therefore N^* will tend to give a closer estimate than $\hat{N}$; therefore N^* is a better estimator than $\hat{N}$.

Two other points of comparison between $\hat{N}$ and N^* are worth noticing.

(i) $\hat{N}$ takes only integral values; N^* can take fractional values. This does not, of course, rule out N^* as an estimator.

(ii) The distribution of $\hat{N}$ is symmetrical, but the distribution of N^* is not. ■

Summarizing the general points about statistical procedures which this example has illustrated, we have seen that a comparison of the usefulness of sampling statistics in providing information about a parameter reduces to a comparison of their sampling distributions. Deliberately, no rules have been provided for this comparison, since what is important in one comparison may be unimportant in another. For example, if we required the sampling distribution of the estimate to be symmetrical about N, then of the two statistics $\hat{N}$ and N^* we would clearly have to choose $\hat{N}$ as the estimator, even though it is not as close to N as N^*.

These remarks should give an idea of the kinds of considerations confronting a statistician. They are not intended to afford a complete study of estimation and of "good" properties which estimates (or rather their sampling distributions) should satisfy. In any event, in some situations we often have to be satisfied with less than the best, because the theory for the best is too intractable.

21.3 THE BINOMIAL DISTRIBUTION

21.3

21.3.0 Introduction

21.3.0

Introduction

There are many situations where an event has a certain probability of occurring, and where we are interested in the number of successes in a sequence of trials. Thus assuming that the probability of a newborn baby being a boy is $\frac{1}{2}$, we could inquire about the probability of any specified number of boys in a family of n. In less down-to-earth vein, we could be interested in the probability of obtaining any specified number of 6's in n throws of a biassed die. Mathematically these two situations are equivalent; in each case we have

$$\left.\begin{aligned} P(\text{"success"}) &= p \\ P(\text{"failure"}) &= 1 - p \end{aligned}\right\} \text{ in a single trial,}$$

and we are interested in the probability of, say, r successes in a sequence of n independent trials.

If X is a random variable corresponding to the number of successes in a single trial, then the probability distribution of X is given by

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

Suppose we take a sample $x_1, \ldots, x_n$ from this distribution. Since each of the x's is either 0 or 1, the total number of successes in this sample is

$$t_n = x_1 + x_2 + \cdots + x_n$$

Now t_n is a realization of a random variable

$$T_n = X_1 + X_2 + \cdots + X_n,$$

so we are interested in the distribution of T_n, which is a sampling statistic. In other words, we are interested once again in a sampling distribution.

21.3.1 Deriving the Distribution

21.3.1

Main Text

Let us begin modestly with $n = 2$.

Denoting a success by S, and a success followed by a failure by SF, etc., we have

$$P(S) = p$$

$$P(F) = 1 - p$$

$$P(SS) = p^2$$

$$P(SF) = p(1 - p)$$

$$P(FS) = (1 - p)p$$

$$P(FF) = (1 - p)^2$$

Therefore,

$$P(T_2 = 0) = P(FF) = (1 - p)^2$$

$$P(T_2 = 1) = P(SF) + P(FS) = 2p(1 - p)$$

$$P(T_2 = 2) = P(SS) = p^2$$

Similarly, for three trials we have

$$P(FFF) = (1-p)^3 \qquad P(FSS) = p^2(1-p)$$
$$P(FFS) = p(1-p)^2 \qquad P(SFS) = p^2(1-p)$$
$$P(FSF) = p(1-p)^2 \qquad P(SSF) = p^2(1-p)$$
$$P(SFF) = p(1-p)^2 \qquad P(SSS) = p^3$$

so that

$$P(T_3 = 0) = (1-p)^3$$
$$P(T_3 = 1) = 3p(1-p)^2$$
$$P(T_3 = 2) = 3p^2(1-p)$$
$$P(T_3 = 3) = p^3$$

It should now be apparent (although it has not been proved) that if a particular sequence contains r successes and $n - r$ failures, then its probability is $p^r(1-p)^{n-r}$. How many sequences are there containing r successes and $n - r$ failures? Let us look at it this way. We may arrive at such a sequence by writing down a string of n circles

$$\bigcirc, \bigcirc, \bigcirc, —, —, —, —, —, \bigcirc$$

and then writing S (corresponding to "success") inside r of the circles and F (corresponding to "failure") inside the remaining $n - r$ circles. For example,

Ⓢ, Ⓢ, Ⓕ, Ⓢ, Ⓕ, . . .

The number of such sequences is the same as the number of ways of selecting r items from n; and this is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Each such sequence has probability $p^r(1-p)^{n-r}$ of occurring. All such sequences are different, that is, exclusive; therefore the probability of getting r successes is equal to the probability of getting one or other of the $\binom{n}{r}$ such sequences, which is

$$\binom{n}{r} p^r(1-p)^{n-r}.$$

Thus we arrive at the probability distribution given by

$$P(T_n = r) = \binom{n}{r} p^r(1-p)^{n-r} \qquad (r = 0, 1, 2, \ldots, n).$$

As we have a distribution, we would expect all the probabilities on the right-hand side to sum to unity. This they do; for representing $1 - p$ by q, we have

$$\begin{aligned} \sum_{r=0}^{n} \binom{n}{r} p^r(1-p)^{n-r} &= \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \\ &= (p+q)^n \qquad \text{(by the binomial theorem)} \\ &= 1 \qquad (p + q = 1) \end{aligned}$$

(See RB9)

Because of the form this distribution takes, we call it the binomial distribution, and we say that T_n is a binomial random variable.

Definition 1

The binomial distribution occurs frequently. Mention has been made earlier of failure in aircraft tyres. In that context we would probably (rather perversely) regard a tyreburst as a "success" and lack of a burst as

"failure". The two-party voting situation may also be looked upon in this way. Those who vote Conservative may be looked on as "successes" (or, of course, "failures"), although it is doubtful whether votes in an actual election may be regarded as a random sample of the potential votes of the whole electorate. On the other hand, a public opinion poll, which selects a few people to find their political views, may obtain a random sample of those views in the electorate. Another example occurs in medical work, where patients treated with a drug will either recover or not. Notice that although the basic model has been described in terms of a *sequence* of trials, the trials may, in fact, be *simultaneous*. For example, a group of hospital patients might all be treated with a drug at the same time.

Exercise 1

Exercise 1
(3 minutes)

(i) Taking $p = \frac{1}{3}$, $n = 6$, work out the complete binomial distribution (i.e. find $P(T_6 = r)$ for $r = 0, 1, \ldots, 6$).

(ii) Find the mean and variance of the distribution

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

(iii) Use the result of (ii) to show that the binomial random variable T_n has

mean $\quad np$

and

variance $\quad np(1 - p)$.

(HINT: Use results (viii) and (ix) in the table on page 29.) ■

Solution 1

Solution 1

(i) $P(T_6 = r) = \binom{n}{r} p^r(1-p)^{n-r}$ where $n = 6$ and $p = \frac{1}{3}$,

$$= \binom{6}{r}\frac{1}{3^r} \times \frac{2^{6-r}}{3^{6-r}}$$

$$= \binom{6}{r}\frac{2^{6-r}}{729}$$

Hence $P(T_6 = 0) = \binom{6}{0}\frac{2^6}{729} = \frac{64}{729}$

$$P(T_6 = 1) = \frac{6 \times 32}{729} = \frac{192}{729}$$

$$P(T_6 = 2) = \frac{15 \times 16}{729} = \frac{240}{729}$$

$$P(T_6 = 3) = \frac{20 \times 8}{729} = \frac{160}{729}$$

$$P(T_6 = 4) = \frac{15 \times 4}{729} = \frac{60}{729}$$

$$P(T_6 = 5) = \frac{6 \times 2}{729} = \frac{12}{729}$$

$$P(T_6 = 6) = \frac{1}{729}$$

(ii) $E(X) = 1 \times p + 0 \times (1-p)$

$= p$

$E(X^2) = 1 \times p + 0 \times (1-p)$

$= p$

Therefore,

$\operatorname{var}(X) = p - p^2$

$= p(1-p)$

(iii) $T_n = X_1 + X_2 + \cdots + X_n$

so that

$$\frac{T_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \overline{X}$$

and we have

$E\left(\frac{T_n}{n}\right) = E(\overline{X}) = E(X)$ (from table)

$= p$ (from (ii)),

whence

$E(T_n) = np$ (see Exercise 21.1.4.3).

Similarly,

$\operatorname{var}\left(\frac{T_n}{n}\right) = \operatorname{var}(\overline{X}) = \sigma^2/n$ (from table)

$= \frac{p(1-p)}{n}$ as $\sigma^2 = \operatorname{var}(X) = p(1-p)$.

Hence

$$\text{var}(T_n) = n^2 \times \frac{p(1-p)}{n} \qquad \text{(see Exercise 21.1.4.3)}$$

$$= np(1-p)$$

■

21.3.2 Estimating an Unknown Probability

21.3.2

Main Text
* *

Our intuitive concepts of probability were derived from the idea of relative frequency in the long run. Hence to estimate the probability of some event, we have merely taken a sequence of independent trials, and worked out the relative frequency of occurrences. We are now in a better position to appreciate what is really happening, and to explain why longer runs tend to give better estimates.

If X is a random variable representing the number (1 or 0) of occurrences of the event in a single trial, then

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

This gives a distribution in which p is a parameter. Therefore, estimating this unknown probability is an example of parametric estimation. Hence we require some suitable sampling statistic.

If x_1 is the number of occurrences on the first trial, x_2 the number on the second, and so forth, the full record of results $x_1, x_2, \ldots, x_n$ constitutes a sample from the distribution of X. The sampling statistic

$$T_n = X_1 + X_2 + \cdots + X_n$$

has mean $E(T_n) = np$ (see Exercise 21.3.1.1), and

$$E\left(\frac{T_n}{n}\right) = p$$

We should therefore accept $\frac{T_n}{n} = \bar{X}$ as a sampling statistic for estimating p, and the realization of $\bar{X}$ in our case is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

In gauging how good an estimator any sampling statistic may be, we look at the sampling variance of the statistic

$$\text{var}(\bar{X}) = \frac{p(1-p)}{n} \qquad \text{(see Exercise 21.3.1.1),}$$

so that as n increases, the variance of the sampling statistic decreases, and the estimator $\bar{X}$ improves.

This establishes the mathematics of the situation. Let us take some numerical examples and see how increases in n affect the sampling distribution of $\frac{T_n}{n} = \bar{X}$.

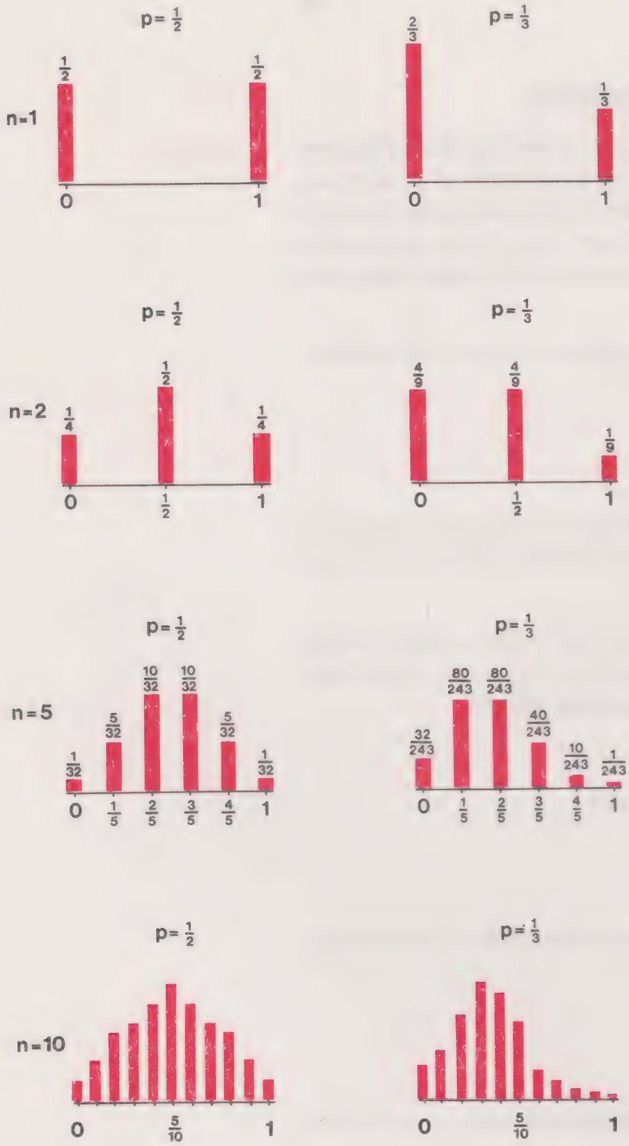
In general, we have

$$P\left(\frac{T_n}{n} = \frac{r}{n}\right) = P(T_n = r)$$

$$= \binom{n}{r} p^r (1-p)^{n-r}$$

$$= p_r \text{ (say)}$$

The following diagrams illustrate the sampling distributions for $n = 1, 2, 5, 10$ in the cases $p = \frac{1}{2}$ and $p = \frac{1}{3}$. They are bar charts of p_r plotted against r/n, where $P(T_n/n = r/n) = p_r$



The important thing to notice about these sampling distributions is the way in which, as the sample size increases, the distribution becomes more and more concentrated about the "true" value of the parameter p. This is a result we expected from our consideration of the variance of $\dfrac{T_n}{n}$.

Exercise 1

(i) Go back to your sequence of 0's and 1's obtained from card guessing in *Unit 16*, and divide up the 500 terms into 50 non-overlapping runs of 10 terms each. For each run, work out the relative frequency of 1's, and then plot these 50 relative frequencies in some way to see how they cluster.

(ii) Repeat this process, but with 20 non-overlapping runs of 25 terms each. Is the clustering any greater? ■

Postscript

"And now sits Expectation in the air."

William Shakespeare
Henry V

M100 – MATHEMATICS FOUNDATION COURSE UNITS

1 Functions
2 Errors and Accuracy
3 Operations and Morphisms
4 Finite Differences
5 NO TEXT
6 Inequalities
7 Sequences and Limits I
8 Computing I
9 Integration I
10 NO TEXT
11 Logic I — Boolean Algebra
12 Differentiation I
13 Integration II
14 Sequences and Limits II
15 Differentiation II
16 Probability and Statistics I
17 Logic II — Proof
18 Probability and Statistics II
19 Relations
20 Computing II
21 Probability and Statistics III
22 Linear Algebra I
23 Linear Algebra II
24 Differential Equations I
25 NO TEXT
26 Linear Algebra III
27 Complex Numbers I
28 Linear Algebra IV
29 Complex Numbers II
30 Groups I
31 Differential Equations II
32 NO TEXT
33 Groups II
34 Number Systems
35 Topology
36 Mathematical Structures